



**República Federativa do Brasil**

Ministério do Desenvolvimento, Indústria,  
Comércio e Serviços

Instituto Nacional da Propriedade Industrial

**(11) BR 112017013524-8 B1**

**(22) Data do Depósito:** 29/12/2015

**(45) Data de Concessão:** 07/03/2023

**(54) Título:** DISPOSITIVO DE COMPUTAÇÃO, MÉTODO E DISPOSITIVO DE ARMAZENAMENTO LEGÍVEL POR COMPUTADOR PARA CONCLUSÃO DE TAREFAS SEM ORIENTAÇÃO DENTRO DE ASSISTENTES PESSOAIS DIGITAIS

**(51) Int.Cl.:** G10L 15/22.

**(30) Prioridade Unionista:** 09/01/2015 US 14/593,584.

**(73) Titular(es):** MICROSOFT TECHNOLOGY LICENSING, LLC.

**(72) Inventor(es):** VISHWAC SENA KANNAN; ALEKSANDAR UZELAC; DANIEL J. HWANG; ROBERT L. CHAMBERS; THOMAS SOEMO; ADINA MAGDALENA TRUFINESCU; KHURAM SHAHID; ALI EMAMI.

**(86) Pedido PCT:** PCT US2015067759 de 29/12/2015

**(87) Publicação PCT:** WO 2016/111881 de 14/07/2016

**(85) Data do Início da Fase Nacional:** 22/06/2017

**(57) Resumo:** A presente invenção refere-se a técnicas para completar uma tarefa de um aplicativo, sem o uso de uma interface gráfica, de um aplicativo no segundo plano de um assistente digital pessoal. Por exemplo, um método pode incluir receber uma entrada de voz por meio de um microfone. O processamento de linguagem natural pode ser realizado usando a entrada de voz para determinar um comando de voz do usuário. O comando de voz do usuário pode incluir uma solicitação para realizar uma tarefa do aplicativo. O aplicativo pode ser levado a executar a tarefa como um processo em segundo plano sem que uma interface de usuário do aplicativo apareça. Uma interface de usuário do assistente digital pessoal pode fornecer uma resposta ao usuário, baseado em um estado recebido associado à tarefa, de modo que a resposta venha de dentro de um contexto da interface de usuário do assistente digital pessoal sem trazer à superfície a interface de usuário do aplicativo.

Relatório Descritivo da Patente de Invenção para  
**"DISPOSITIVO DE COMPUTAÇÃO, MÉTODO E DISPOSITIVO DE  
ARMAZENAMENTO LEGÍVEL POR COMPUTADOR PARA  
CONCLUSÃO DE TAREFAS SEM ORIENTAÇÃO DENTRO DE  
ASSISTENTES PESSOAIS DIGITAIS".**

ANTECEDENTES DA INVENÇÃO

[001] Com os avanços da tecnologia de computação, surgiram dispositivos de computação cada vez mais poderosos. Por exemplo, cada vez mais os dispositivos de computação estão adicionando recursos como o reconhecimento de fala. A fala pode ser uma forma eficaz para um usuário se comunicar com um dispositivo de computação, e aplicativos controlados por fala estão sendo desenvolvidos, tais como assistentes digitais pessoais controlados por fala.

[002] Um assistente digital pessoal pode ser usado para realizar tarefas ou serviço para um indivíduo. Por exemplo, o assistente pessoal digital pode ser um módulo de software em execução em um dispositivo móvel ou em um computador de mesa. Exemplos de tarefas e serviços que podem ser realizados pelo assistente digital pessoal podem incluir a obtenção de condições e previsões climáticas, resultados de partidas, rotas e condições de trânsito, notícias regionais e/ou nacionais, e preços das ações; gerenciar a agenda do usuário pela criação de novos itens da agenda, e lembrar o usuário dos eventos próximos; e armazenar e recuperar lembretes.

[003] No entanto, é provável que o assistente digital pessoal não consiga realizar todas as tarefas que um usuário pode desejar ter realizado. Portanto, pode existir ampla oportunidade para aperfeiçoamento nas tecnologias relacionadas aos assistentes digitais pessoais controlados por fala.

SUMÁRIO

[004] A intenção deste sumário é a de apresentar, de maneira simplificada, uma seleção de conceitos descritos em detalhes a seguir na Descrição Detalhada. O presente Sumário não tem a intenção de identificar aspectos cruciais ou essenciais da matéria reivindicada e não deve ser usado para limitar o âmbito da matéria reivindicada.

[005] São descritas técnicas e ferramentas para completar uma tarefa de um aplicativo, sem orientação (*headless*), de um aplicativo no segundo plano de um assistente digital pessoal. Por exemplo, um método pode ser implementado por um dispositivo de computação compreendendo um microfone. O método pode compreender receber, por um assistente digital pessoal controlado por voz, uma entrada de voz digital gerada por um usuário. A entrada de voz digital pode ser recebida por meio do microfone. O processamento de linguagem natural pode ser realizado usando a entrada de voz digital para determinar um comando de voz do usuário. O comando de voz do usuário pode compreender uma solicitação para realizar uma função predefinida de um aplicativo de terceiros com suporte a voz. A função predefinida pode ser identificada usando uma estrutura de dados que define funções suportadas pelos aplicativos de terceiros com suporte a voz disponíveis usando a entrada de voz. O aplicativo de terceiros com suporte a voz pode ser levado a executar a função predefinida como um processo em segundo plano sem uma interface do usuário do aplicativo de terceiros com suporte a voz aparecendo em um meio de exibição do dispositivo de computação. Uma resposta pode ser recebida do aplicativo de terceiros com suporte a voz, indicando um estado associado à função predefinida. Uma interface de usuário do assistente digital pessoal controlado por voz pode fornecer uma resposta ao usuário, baseado no estado recebido associado à função predefinida, de modo que a resposta venha de dentro de um contexto da interface de usuário do assistente digital pessoal controlado por voz sem trazer para a

superfície a interface de usuário do aplicativo de terceiros com suporte a voz.

[006] Como outro exemplo, dispositivos de computação compreendendo unidades de processamento, memória e um ou mais microfones podem ser proporcionados para realizar as operações descritas aqui. Por exemplo, um método realizado pelo dispositivo de computação pode incluir receber a entrada de fala gerada por um usuário por meio de um ou mais microfones. O reconhecimento de fala pode ser realizado usando a entrada de fala para determinar um comando falado. O comando falado pode compreender uma solicitação para realizar uma tarefa de um aplicativo de terceiros. A tarefa pode ser identificada usando uma estrutura de dados que define as tarefas dos aplicativos de terceiros que podem ser invocados pelo comando falado. Pode ser determinado se a tarefa do aplicativo de terceiros é capaz de ser executada sem orientação. O aplicativo de terceiros pode ser executado como um processo em segundo plano para executar a tarefa sem orientação quando for determinado que a tarefa do aplicativo de terceiros é capaz de ser executada sem orientação. Uma resposta do aplicativo de terceiros pode ser recebida indicando um estado associado à tarefa. Uma interface de usuário do assistente digital pessoal controlado por fala pode fornecer uma resposta ao usuário, baseado no estado recebido associado à tarefa, de modo que a resposta venha de dentro de um contexto da interface de usuário do assistente digital pessoal controlado por fala sem trazer para a superfície a interface de usuário do aplicativo de terceiros.

[007] Como outro exemplo, dispositivos de computação compreendendo unidades de processamento e memória podem ser proporcionados para realizar as operações descritas aqui. Por exemplo, um dispositivo de computação pode realizar operações para completar uma tarefa de um aplicativo com suporte a voz dentro do contexto de



um assistente digital pessoal controlado por voz. As operações podem compreender receber uma entrada de voz digital gerada por um usuário no assistente digital pessoal controlado por voz. A entrada de voz digital pode ser recebida por meio de um microfone. O processamento de linguagem natural pode ser realizado usando a entrada de voz digital para determinar um comando de voz do usuário. O comando de voz do usuário pode compreender uma solicitação para realizar a tarefa do aplicativo com suporte a voz. A tarefa pode ser identificada usando uma estrutura de dados extensível que mapeia os comandos de voz do usuário para as tarefas dos aplicativos com suporte a voz. Pode ser determinado se a tarefa do aplicativo com suporte a voz é uma tarefa de primeiro plano ou uma tarefa de segundo plano. Quando for determinado que a tarefa é uma tarefa de segundo plano, o aplicativo com suporte a voz pode ser levado a executar a tarefa como uma tarefa de segundo plano e dentro de um contexto do assistente digital pessoal controlado por voz sem uma interface de usuário do aplicativo com suporte a voz ser trazida para a superfície. Uma resposta do aplicativo com suporte a voz pode ser recebida. A resposta pode indicar um estado associado à tarefa. Uma resposta pode ser fornecida ao usuário baseado no estado recebido associado à tarefa. A resposta pode ser fornecida dentro do contexto do assistente digital pessoal controlado por voz sem uma interface de usuário do aplicativo com suporte a voz ser trazida para a superfície quando for determinado que a tarefa é uma tarefa de segundo plano.

[008] Como descrito aqui, uma variedade de outros aspectos e vantagens podem ser incorporados às tecnologias conforme desejado.

#### BREVE DESCRIÇÃO DOS DESENHOS

[009] A Figura 1 é um diagrama representando um exemplo de um sistema para completar sem orientação uma tarefa de um aplicativo no segundo plano de um assistente digital pessoal.

[0010] A Figura 2 é um diagrama representando um exemplo de arquitetura de software para completar sem orientação uma tarefa de um aplicativo um diagrama representando um exemplo de arquitetura de software no segundo plano de um assistente digital pessoal.

[0011] A Figura 3 é um diagrama de um exemplo de máquina de estado para um aplicativo fazendo interface com um assistente digital pessoal.

[0012] A Figura 4 é um exemplo de uma definição de comando que pode ser usada para criar uma estrutura de dados para ativar uma interface entre um aplicativo e um assistente digital pessoal.

[0013] A Figura 5 é um exemplo de diagrama de sequência ilustrando a comunicação de múltiplos encadeamentos para realizar uma tarefa de um aplicativo, sem orientação, de dentro de um assistente digital pessoal.

[0014] A Figura 6 é um fluxograma de um exemplo de método para completar sem orientação uma tarefa de um aplicativo, no segundo plano de um assistente digital pessoal.

[0015] A Figura 7 é um fluxograma de um método ilustrativo para determinar se um aplicativo deverá ser pré-carregado enquanto um usuário está falando para um assistente digital pessoal.

[0016] A Figura 8 é um diagrama de um exemplo de sistema de computação no qual algumas concretizações descritas podem ser implementadas.

[0017] A Figura 9 é um exemplo de dispositivo móvel que pode ser usado em conjunto com as tecnologias descritas aqui.

[0018] A Figura 10 é um exemplo de ambiente com suporte a nuvem que pode ser usado em conjunto com as tecnologias descritas aqui.

## DESCRIÇÃO DETALHADA

### Visão Geral

[0019] À medida que o usuário se familiariza com o uso do assistente digital pessoal, o usuário pode preferir realizar mais ações dentro do contexto o assistente digital pessoal. No entanto, o fornecedor do assistente digital pessoal não pode prever ou investir tempo no desenvolvimento de todos os aplicativos que um usuário pode desejar utilizar. Assim, pode ser desejável que o assistente digital pessoal seja capaz de chamar ou inicializar aplicativos de terceiros que são criados por outras entidades que não o fornecedor do assistente digital pessoal.

[0020] Em uma solução típica, a interface do usuário do aplicativo é trazida para a superfície quando o assistente digital pessoal inicia o aplicativo e o controle do programa passa do assistente digital pessoal para o aplicativo. Após a interface de usuário do aplicativo ser trazida para a superfície, o usuário pode verificar a condição da solicitação e o usuário pode realizar tarefas adicionais de dentro do aplicativo. Para retornar para a interface de usuário do assistente digital pessoal, o usuário deverá sair do aplicativo antes de o controle poder ser retornado ao assistente digital pessoal.

[0021] Como um exemplo específico de uso de um assistente digital pessoal de um telefone móvel, o usuário pode solicitar que um filme seja adicionado à fila do usuário usando um aplicativo de filme instalado no telefone móvel. Por exemplo, o usuário pode dizer "Aplicativo de Filme, adicione o Filme X a minha fila" para a interface do usuário do assistente digital pessoal. Após o comando ser falado e reconhecido pelo assistente, o assistente pode iniciar o aplicativo de filme, que irá apresentar a interface do usuário do aplicativo de filme. O filme pode ser adicionado à fila do usuário e a fila pode ser apresentada ao usuário como a verificação de que o filme foi adicionado. O usuário pode

continuar a usar o aplicativo de filme ou o usuário pode fechar o aplicativo de filme para retornar para a interface de usuário do assistente digital pessoal.

[0022] Quando um assistente digital pessoal realiza a transição do controle para o aplicativo, o carregamento do aplicativo e de sua interface de usuário para a memória pode consumir uma quantidade de tempo perceptível. O atraso pode potencialmente impactar a produtividade do usuário, tal como gerando atraso para o usuário ao realizar uma tarefa seguinte e/ou interrompendo a linha de raciocínio do usuário. Por exemplo, a atenção do usuário pode ser direcionada ao fechamento do aplicativo antes de retornar para a interface do usuário do assistente digital pessoal. Adicionalmente, por realizar a transição do controle para o aplicativo, as informações contextuais disponíveis ao assistente digital pessoal podem não estar disponíveis ao aplicativo. Por exemplo, o assistente digital pessoal pode entender a identidade e informações de contato do cônjuge do usuário, a localização da residência ou escritório do usuário, ou a localização de um fornecedor de serviços de creche do usuário, mas o aplicativo pode não ter acesso às informações contextuais.

[0023] Nas técnicas e soluções descritas aqui, um assistente digital pessoal pode determinar se uma tarefa de um aplicativo de terceiros pode ser realizada no segundo plano, de modo que as operações para realizar a tarefa sejam realizadas dentro do contexto do assistente digital pessoal e sem uma interface de usuário do aplicativo com suporte a voz ser trazida para a superfície. Dessa forma, o usuário pode experimentar que um dado conjunto de tarefas seja realizado dentro do contexto do assistente digital pessoal, em vez do contexto do aplicativo que está realizando a tarefa do usuário. Adicionalmente, o consumo de energia do dispositivo pode potencialmente ser reduzido (e a vida útil da bateria prolongada), uma vez que a interface do usuário do aplicativo

não é carregada na memória quando a tarefa do aplicativo é realizada no segundo plano.

[0024] Os aplicativos podem se registrar no assistente digital pessoal para expandir a lista de capacidades nativas que o assistente oferece. Os aplicativos podem ser instalados em um dispositivo ou chamados através de uma rede (tal como a Internet) como um serviço. Uma definição de esquema pode permitir que os aplicativos registrem um comando de voz com uma solicitação para serem inicializados sem orientação quando um usuário solicita esse comando/tarefa. Por exemplo, os aplicativos podem incluir um arquivo de definição de comando de voz (VCD) acessível pelo assistente digital pessoal, em que o arquivo VCD identifica tarefas que podem ser inicializadas sem orientação. A definição pode especificar que a tarefa do aplicativo é sempre ser inicializado sem orientação, ou a definição pode especificar que a tarefa do aplicativo é para ser inicializada sem orientação sob circunstâncias específicas. Por exemplo, um aplicativo poderia escolher fazer algo sem orientação se o usuário estiver solicitando que a tarefa seja realizada em um dispositivo que não possui uma superfície de exibição (tal como uma pulseira esportiva sem fio), ou quando o usuário estiver operando no modo viva-voz (tal como quando o usuário está conectado a um fone de cabeça Bluetooth).

[0025] Os aplicativos podem fornecer uma resposta a respeito do progresso, falha e êxito da conclusão da tarefa solicitada e a saída relacionada aos estados pode ser fornecida pela interface do usuário do assistente digital pessoal. Os aplicativos podem fornecer muitos tipos diferentes de dados de volta ao assistente digital pessoal, incluindo texto de exibição, texto que pode ser lido em voz alta, uma ligação profunda (*deep link*) de volta ao aplicativo, uma ligação para uma página da Rede ou site da Rede, e conteúdo da Rede baseado na Linguagem de Marcação de Hipertexto (HTML), por exemplo. Os dados do aplicativo

para o assistente podem ser apresentados como se estivessem vindo de uma função nativa do assistente por meio da interface de usuário do assistente.

[0026] Se o usuário fornecer uma solicitação para o aplicativo que possa ter múltiplos significados ou resultados, o aplicativo pode fornecer ao assistente digital pessoal uma lista de escolhas e a interface do usuário do assistente pode ser usada para desambiguar entre as escolhas. Se o usuário fornecer uma solicitação para o aplicativo que pode ser destrutiva ou que é importante (tal como quando o usuário solicita que um aplicativo de banco realize uma transferência de saldos), uma interface de confirmação do assistente pode ser usada para confirmar a solicitação antes de completar a tarefa importante ou destrutiva.

[0027] Os aplicativos podem ser carregados ou pré-carregados de forma especulativa conforme os comandos são falados. Por exemplo, quando o usuário completa a frase "Aplicativo de Filme" a partir do comando "Aplicativo de Filme, adicione Filme X à minha fila", a memória pode ser alocada, e várias sub-rotinas de um aplicativo de filme instalado podem ser recuperadas a partir do armazenamento e carregadas para a memória alocada em preparação para utilizar as sub-rotinas quando o comando estiver completo. Quando o aplicativo é um serviço da Rede, o pré-carregamento pode incluir estabelecer uma sessão de comunicação e recuperar informações específicas do usuário a partir de uma base de dados em um servidor remoto, por exemplo. Por pré-carregar o aplicativo, o tempo para responder para o usuário pode potencialmente ser reduzido de modo que a interação seja mais natural e de modo que o usuário possa mover para a próxima tarefa mais rapidamente, tornando o usuário mais produtivo.

[0028] Usando as tecnologias reveladas aqui, o usuário que deseja adicionar um filme à fila do usuário com um aplicativo de filme pode ter

uma experiência diferente de quando utiliza a solução típica de inicializar o aplicativo de filme e passar o controle para o aplicativo. Neste exemplo, o comando "adicionar filme à fila do Aplicativo de Filme" pode ser definido como sem orientação em uma estrutura de dados de comando, tal como um arquivo VCD. Quando o usuário diz "Aplicativo de Filme" a partir do comando, "Aplicativo de Filme, adicione o Filme X à minha fila", o aplicativo de filme pode ser pré-carregado de modo que o tempo de resposta para o usuário possa ser reduzido. Quando o comando estiver completo, o filme pode ser adicionado à fila do usuário usando o aplicativo de filme, mas sem trazer para a superfície da interface de usuário do aplicativo de filme. O filme pode ser adicionado à fila do usuário e o assistente digital pessoal pode confirmar (usando a interface de usuário do assistente) que o filme foi adicionado. O usuário pode experimentar um tempo de resposta mais rápido e pode realizar menos etapas para completar a tarefa (por exemplo, o aplicativo de filme não precisa ser fechado).

### **Exemplo de Sistema incluindo um Assistente Digital Pessoal**

[0029] A Figura 1 é um diagrama de sistema representando um exemplo de um sistema 100 para completar, sem orientação, uma tarefa 112 de um aplicativo com suporte a voz 110 no segundo plano de um assistente digital pessoal 120. O aplicativo com suporte a voz 110 e o assistente digital pessoal 120 podem ser módulos de softwares instalados em um dispositivo de computação 130. O dispositivo de computação 130 pode ser um computador de mesa, um laptop, um telefone móvel, um smartphone, um dispositivo portátil (tal como um relógio de pulso ou uma pulseira eletrônica sem fio), ou um computador tablet, por exemplo. O dispositivo de computação 130 pode incluir uma estrutura de dados de comando 140 para identificar aplicativos e tarefas de aplicativos que podem ser inicializados pelo assistente digital pessoal 120. Os aplicativos podem ser inicializados pelo assistente pessoal

digital 120 no primeiro plano (tal como onde uma interface do usuário do aplicativo aparece quando o aplicativo é inicializado) e/ou no segundo plano (tal como onde a interface do usuário do aplicativo não aparece quando o aplicativo é inicializado). Por exemplo, algumas tarefas de um aplicativo podem ser inicializadas no segundo plano e diferentes tarefas do mesmo aplicativo podem ser inicializadas no segundo plano. A estrutura de dados de comando 140 pode definir de que maneira o aplicativo e/ou as tarefas do aplicativo devem ser inicializadas a partir do assistente digital pessoal 120.

[0030] O dispositivo de computação 130 pode incluir um microfone 150 para converter o som em um sinal elétrico. O microfone 150 pode ser um microfone dinâmico, condensador ou piezelétrico usando indução eletromagnética, uma alteração na capacitância, ou piezeletricidade, respectivamente, para produzir o sinal elétrico a partir das variações de pressão de ar. O microfone 150 pode incluir um amplificador, um ou mais filtros analógicos ou digitais e/ou um conversor de analógico para digital para produzir uma entrada de som digital. A entrada de som digital pode compreender uma reprodução da voz do usuário, tal como quando o usuário está comandando o assistente digital pessoal 120 para realizar uma tarefa. O dispositivo de computação 130 pode incluir uma tela de toque ou teclado (não ilustrado) para permitir que o usuário insira entrada textual.

[0031] A entrada de som digital e/ou a entrada textual pode ser processada por um módulo de processamento de linguagem natural 122 do assistente digital pessoal 120. Por exemplo, o módulo de processamento de linguagem natural 122 pode receber a entrada de som digital e traduzir as palavras faladas por um usuário em texto. O texto extraído pode ser analisado semanticamente para determinar um comando de voz do usuário. Por analisar a entrada de som digital e realizar ações em resposta aos comandos falados, o assistente pessoal



digital 120 pode ser controlado por voz. Por exemplo, o assistente digital pessoal 120 pode comparar o texto extraído com uma lista de possíveis comandos do usuário para determinar o comando que mais provavelmente corresponde à intenção do usuário. A correlação pode ser baseada em métodos estatísticos ou probabilísticos, árvores de decisão ou em outras regras, em outros critérios de correlação apropriados, ou combinações dos mesmos. Os comandos em potencial do usuário podem ser comandos nativos do assistente digital pessoal 120 e/ou comandos definidos na estrutura de dados de comando 140. Assim, por definir comandos na estrutura de dados de comando 140, a variedade de tarefas que podem ser realizadas a favor do usuário pelo assistente digital pessoal 120 pode ser estendida. Os comandos potenciais podem incluir realizar a tarefa 112 do aplicativo com suporte a voz 110, que pode ser definida como sendo uma tarefa de segundo ou plano ou sem orientação na estrutura de dados de comando 140.

[0032] O módulo de processamento de linguagem natural 122 pode gerar um fluxo de texto à medida que a fala é processada, de modo que cadeias de caracteres de texto intermediárias possam ser analisadas antes de uma expressão vocal do usuário estar completa. Assim, se o usuário iniciar um comando com um nome de um aplicativo, o aplicativo pode ser identificado cedo na expressão vocal, e o aplicativo pode ser pré-carregado antes de o usuário completar o comando. O pré-carregamento do aplicativo pode incluir recuperar instruções do aplicativo a partir de uma memória não volátil relativamente mais lenta (tal como uma unidade de disco rígido ou memória Flash) e armazenar as instruções em uma memória volátil relativamente mais rápida (tal como memória principal ou memória cache).

[0033] Quando o assistente digital pessoal 120 determina que um comando está associado a uma tarefa de um aplicativo, a tarefa do aplicativo pode ser executada. Se o assistente digital pessoal 120

determinar que a tarefa do aplicativo deverá ser executada com um processo em segundo plano (tal como pela análise da definição na estrutura de dados de comando 140), o aplicativo pode ser executado no segundo plano. O aplicativo, tal como o aplicativo com suporte a voz 110, pode se comunicar com o assistente digital pessoal 120. Por exemplo, o aplicativo pode sequenciar através de um conjunto de estados associados à conclusão da tarefa, e o estado do aplicativo pode ser comunicado ao assistente digital pessoal 120. Por exemplo, o aplicativo pode começar em um estado "inicial", transita para um estado "em progresso" enquanto a tarefa está sendo realizada, e então transita para um estado "final" quando a tarefa está completa.

[0034] O assistente digital pessoal 120 pode relatar o progresso da tarefa por meio de uma interface do usuário 124. A interface do usuário 124 pode comunicar informações para o usuário de diversas maneiras, tal como pela apresentação de texto, gráficos ou hiperligações em um meio de exibição do dispositivo de computação 130, gerando saídas de áudio a partir de um alto-falante do dispositivo de computação 130, ou gerando outras saídas sensoriais, tais como vibrações a partir de um motor elétrico conectado a um peso excêntrico do dispositivo de computação 130. Por exemplo, a interface do usuário 124 pode fazer com que uma roda giratória seja apresentada em uma tela de exibição do dispositivo de computação 130 quando a tarefa estiver no estado em progresso. Como outro exemplo, a interface de usuário 124 pode gerar fala simulada, indicando a conclusão bem-sucedida da tarefa quando a tarefa está no estado final e a tarefa tiver sido completada com sucesso. Usando a interface de usuário 124 do assistente digital pessoal 120 para relatar a condição da tarefa, a resposta pode vir de dentro de um contexto da interface do usuário 124 sem trazer uma interface de usuário do aplicativo para a superfície.

[0035] Deve-se observar que o aplicativo com suporte a voz 110

pode ser criado pelo produtor do assistente digital pessoal 120 ou por uma entidade externa que é diferente do produtor. A interoperação do assistente digital pessoal 120 e da aplicação com suporte a voz 110 pode ser alcançada mediante comprimento de um contrato de software de aplicativo para aplicativo e definindo funcionalidade na estrutura de dados de comando 140. O aplicativo com suporte a voz 110 pode ser capaz de operar como um aplicativo independente ou somente como um componente do assistente digital pessoal 120. Como um aplicativo independente, o aplicativo com suporte a voz 110 pode ser inicializado fora do assistente pessoal digital 120 como um processo de segundo plano, tal como por meio de toque ou clique duplo em um ícone associado ao aplicativo com suporte a voz 110 e exibido em uma tela de exibição do dispositivo de computação 130. O aplicativo com suporte a voz 110 pode apresentar uma interface do usuário quando é inicializado e o usuário pode interagir com a interface do usuário para realizar tarefas. A interação pode ser somente com entrada de voz, ou outros modos de entrada também podem ser usados, tal como entrada de texto ou gestos. Os aplicativos chamados pelo assistente digital pessoal 120 podem ser instalados no dispositivo de computação 130 ou podem ser serviços da Rede.

[0036] O assistente digital pessoal 120 pode chamar serviços da Rede, tal como o serviço da Rede 162 em execução no computador servidor remoto 160. Os serviços da Rede são funções de software proporcionadas em um endereço de rede através de uma rede, tal como a rede 170. A rede 170 pode incluir uma rede local (LAN), uma Rede de Longa Distância (WAN), a Internet, uma intranet, uma rede com fio, uma rede sem fio, uma rede celular, combinações dos mesmos, ou qualquer rede adequada para proporcionar um canal para comunicação entre o dispositivo de computação 130 e o computador servidor remoto 160. Deve-se apreciar que a topologia de rede ilustrada na Figura 1 foi

simplificada e que múltiplas redes e dispositivos de rede podem ser utilizados para interconectar os vários sistemas de computação revelados aqui. O serviço da Rede 162 pode ser chamado como parte do núcleo ou como parte do assistente digital pessoal 120. Por exemplo, o serviço da Rede 162 pode ser chamado como uma sub-rotina do módulo de processamento de linguagem natural 122. Adicionalmente ou como alternativa, o serviço da Rede 162 pode ser um aplicativo definido na estrutura de dados de comando 140 e pode ser capaz de ser inicializado sem orientação a partir do assistente digital pessoal 120.

### **Exemplo de Arquitetura de Software incluindo um Assistente Digital Pessoal**

[0037] A Figura 2 é um diagrama representando um exemplo de arquitetura de software 200 para completar uma tarefa de um aplicativo sem orientação no segundo plano de um assistente digital pessoal 120. Ao realizar uma tarefa de um aplicativo sem orientação, a tarefa pode ser executada no segundo plano e uma interface de usuário do aplicativo não é trazida para a superfície como resultado da tarefa sendo realizada. Em vez disso, a interface de usuário do assistente digital pessoal 120 pode ser usada para fornecer saída para e/ou entrada a partir do usuário de modo que o usuário interaja dentro do contexto do assistente pessoal digital 120 e não no contexto do aplicativo. Assim, uma tarefa executada sem orientação de um aplicativo pode ser executada no segundo plano pela duração da execução da tarefa, e a interface de usuário do aplicativo nunca é trazida à superfície. Um dispositivo de computação, tal como o dispositivo de computação 130, pode executar software para um assistente digital pessoal 120, um núcleo 210 do sistema operacional (SO), e um aplicativo 230 organizado de acordo com a arquitetura 200.

[0038] O núcleo do SO 210 geralmente proporciona uma interface entre os componentes de software e os componentes de hardware do

dispositivo de computação 130. O núcleo do SO 210 pode incluir componentes para sintetização (por exemplo, sintetização de saída visual a um meio de exibição, geração de saída de voz e outros sons para um alto-falante, e geração de uma saída de vibração para um motor elétrico), componentes para ligação à rede, componentes para gerenciamento de processo, componentes para gerenciamento de memória, componentes para rastreamento de localização, e componentes para reconhecimento de fala e outros processamentos de entrada. O núcleo do SO 210 pode gerenciar as funções de entrada do usuário, funções de saída, funções de acesso ao armazenamento, funções de comunicação de rede, funções de gerenciamento de memória, funções de gerenciamento de processos e outras funções para o dispositivo de computação 130. O núcleo do SO 210 pode fornecer acesso a tais funções ao assistente digital pessoal 120 e ao aplicativo 230, tal como através de várias chamadas do sistema.

[0039] Um usuário pode gerar entrada do usuário (tal como voz, entrada tátil e movimento) para interagir com o assistente digital pessoal 120. O assistente digital pessoal 120 pode ser conscientizado da entrada do usuário por meio do núcleo do SO 210, que pode incluir funcionalidade para criar mensagens em resposta à entrada do usuário. As mensagens podem ser usadas pelo assistente digital pessoal 120 ou por outro software. A entrada do usuário pode incluir entrada tátil, tal como entrada de tela de toque, pressionamentos de botão ou pressionamentos de tecla. O núcleo do SO 210 pode incluir funcionalidade para reconhecer toques, gestos com os dedos, etc. em uma tela de toque a partir da entrada tátil, entrada de botão ou entrada de pressionamento de tecla. O núcleo do SO 210 pode receber entrada a partir do microfone 150 e pode incluir funcionalidade para reconhecer comandos e/ou palavras faladas a partir da entrada de voz. O núcleo do SO 210 pode receber entrada de um acelerômetro e pode incluir

funcionalidade para reconhecer orientação ou movimento, tal como movimento de agitar.

[0040] O mecanismo de processamento de entrada 222 da interface do usuário (IU) do assistente digital pessoal 120 pode aguardar por mensagens de eventos de entrada do usuário a partir do núcleo do SO 210. As mensagens de eventos da IU podem indicar uma palavra reconhecida a partir da entrada de voz, um gesto de "panning" (usar um ou dois dedos para arrastar horizontalmente a tela), um gesto de movimento rápido (flicking), um gesto de arrastar, ou outro gesto em uma tela de toque do dispositivo, um leve toque na tela de toque, entrada de pressionamento de tecla, um gesto de sacudir ou outro evento da IU (por exemplo, botões direcionais ou entrada de dispositivo *trackball*). O mecanismo de processamento de entrada da IU 222 pode traduzir as mensagens de eventos da IU a partir do núcleo do SO 210 em informações enviadas à lógica de controle 224 do assistente digital pessoal 120. Por exemplo, o mecanismo de processamento de entrada da IU 222 pode incluir capacidades de processamento de linguagem natural e pode indicar que um determinado nome de aplicativo foi falado ou digitado ou que um comando de voz foi fornecido pelo usuário. Como alternativa, as capacidades de processamento de linguagem natural podem ser incluídas na lógica de controle 224.

[0041] A lógica de controle 224 pode receber informações de vários módulos do assistente pessoal digital 120, tal como o mecanismo de processamento de entrada da IU 222, um repositório de informações personalizado 226, e a estrutura de dados de comando 140, e a lógica de controle 224 podem efetuar decisões e realizar operações baseado nas informações recebidas. Por exemplo, a lógica de controle 224 pode determinar se o assistente digital pessoal 120 deve realizar uma tarefa em nome do usuário, tal como analisar um fluxo de texto falado para determinar se um comando de voz foi informado.

[0042] A lógica de controle 224 pode aguardar até que todo o comando do usuário seja falado antes de agir sobre o comando, ou a lógica de controle 224 pode começar a agir sobre o comando à medida que ainda está sendo falado e antes de ser completado. Por exemplo, a lógica de controle 224 pode analisar cadeias de caracteres intermediárias do comando falado e tentar correlacionar as cadeias de caracteres com um ou mais aplicativos definidos na estrutura de dados de comando 140. Quando a probabilidade de que um aplicativo será chamado excede um limiar, o aplicativo pode ser pré-carregado de modo que o aplicativo possa responder ao usuário mais prontamente. Múltiplos aplicativos e/ou funções podem ser especulativamente pré-carregados antes de serem chamados, e os aplicativos podem ser interrompidos caso seja determinado que o aplicativo não será chamado. Por exemplo, quando o usuário inicia o comando falado com o nome de um aplicativo específico, há uma grande probabilidade de que o aplicativo específico será chamado, e, portanto, este aplicativo pode ser pré-carregado. Como outro exemplo, algumas cadeias de caracteres de comando parcial podem ser limitadas a um pequeno conjunto de aplicativos definidos na estrutura de dados de comando 140, e o conjunto de aplicativos pode ser pré-carregado em paralelo quando houver uma correlação na cadeia de caracteres de comando parcial. Especificamente, a estrutura de dados de comando 140 pode ter apenas dois aplicativos com comandos contendo a palavra "fazer", tal como um aplicativo de câmera com um comando "fazer uma foto" e um aplicativo de anotações com um comando "fazer uma anotação". A lógica de controle 224 pode começar a pré-carregar tanto o aplicativo de câmera quanto o aplicativo de anotações quando a palavra "tirar" for reconhecida, e então o aplicativo de anotações pode ser interrompido quando o comando completo "tirar uma foto" for reconhecido. O pré-carregamento do aplicativo pode incluir alocar memória, realizar a busca

antecipada (pre-fetching) de instruções, estabelecer uma sessão de comunicação, recuperar informações de um banco de dados, iniciar um novo encadeamento de execução (thread), acionar uma interrupção ou outras operações adequadas específicas ao aplicativo. Os serviços do núcleo do SO 210 podem ser chamados durante o pré-carregamento, tal como o serviço de gerenciamento de processos, o serviço de gerenciamento de memória e o serviço de rede, por exemplo.

[0043] O texto falado pode incluir informações contextuais e a lógica de controle 224 pode resolver a informação contextual de modo que o comando de voz do usuário esteja livre de contexto. As informações contextuais podem incluir uma localização atual, uma hora atual, uma orientação do dispositivo de computação 130, e informações pessoais armazenadas no repositório de informações personalizado 226. As informações pessoais podem incluir: relacionamentos do usuário, tal como o nome de um usuário, de um cônjuge ou filho; localizações específicas do usuário, tal como endereço de residência, trabalho, escola, creche ou médico; informações da lista de contatos ou calendário do usuário; a cor restaurante ou método de transporte favorito do usuário; aniversários importantes, datas comemorativas ou outras datas; e outras informações específicas do usuário. O usuário pode fornecer um comando com informações textuais e a lógica de controle 224 pode traduzir o comando em um comando livre de contexto. Por exemplo, o usuário pode fornecer o comando "Aplicativo de ônibus, diga os ônibus que me levam para casa dentro da próxima hora". Neste exemplo, as informações contextuais no comando são a data e hora atual, a localização atual e a localização da residência do usuário.

[0044] A lógica de controle 224 pode obter a hora atual pelo núcleo do SO 210, que pode manter ou ter acesso a um relógio em tempo real. A lógica de controle 224 pode obter dados de localização atual para o



dispositivo de computação 130 a partir do núcleo do SO 210, que pode obter os dados de localização atuais a partir de um componente local do dispositivo de computação 130. Por exemplo, os dados de localização podem ser determinados com base nos dados a partir de um sistema de posicionamento global (GPS), por triangulação entre as torres de uma rede celular, por referência às localizações físicas ou roteadores Wi-Fi na vizinhança, ou por outro mecanismo. A lógica de controle 224 pode obter a localização da residência do usuário a partir do repositório de informações personalizado 226. O repositório de informações personalizado 226 pode ser armazenado em armazenamento auxiliar ou em outro armazenamento não volátil do dispositivo de computação 130. Assim, a lógica de controle 224 pode receber as informações personalizadas por meio do núcleo SO 210 que pode acessar o recurso de armazenamento (por exemplo, o repositório de informações personalizado 226). Quando as informações contextuais podem ser resolvidas, o comando pode ser traduzido em um comando livre de contexto. Por exemplo, se for Sexta-Feira às 6:00 da tarde, o usuário estiver na 444 Main Street, e a residência do usuário for 123 Pleasant Drive, então o comando livre de contexto pode ser "Aplicativo de ônibus, informe os ônibus que passam próximo à 444 Main Street e que passam próximo às 123 Pleasant Drive entre 6:00 e 7:00 da tarde às Sextas-Feiras".

[0045] O comando do usuário pode ser realizado pela lógica de controle 224 (tal como quando o comando é um comando nativo do assistente digital pessoal 120), um aplicativo 230 instalado no dispositivo de computação 130 (tal como quando o comando é associado ao aplicativo 230), ou o serviço da Rede 162 (tal como quando o comando é associado ao serviço da Rede 162). A estrutura de dados de comando 140 pode especificar quais comandos estão associados a quais aplicativos e se o comando pode ser realizado no

primeiro plano ou no segundo plano. Por exemplo, a estrutura de dados de comando 140 pode mapear os comandos de voz do usuário para funções suportadas por aplicativos de terceiros com suporte a voz disponíveis.

[0046] A lógica de controle 224 pode fazer com que uma função predefinida 232 do aplicativo 230 seja executada quando a lógica de controle 224 determinar que o comando do usuário está associado à função predefinida 232 do aplicativo 230. Se a lógica de controle 224 determinar que a função predefinida 232 do aplicativo 230 deve ser executada como um processo de segundo plano, a função predefinida 232 pode ser executada no segundo plano. Por exemplo, a lógica de controle 224 pode enviar uma solicitação 240 à função predefinida 232 acionando uma interrupção, gravando na memória compartilhada, gravando em uma fila de mensagens, passando uma mensagem, ou iniciando um novo encadeamento de execução (tal como por meio do componente de gerenciamento de processo do núcleo SO 210). O aplicativo 230 pode realizar a função predefinida 232 e retornar uma resposta 242 à lógica de controle 224 acionando uma interrupção, gravando na memória compartilhada, gravando em uma fila de mensagem ou passando uma mensagem. A resposta pode incluir um estado do aplicativo 230 e/ou outras informações em resposta ao comando do usuário.

[0047] A lógica de controle 224 pode fazer com que o serviço da Rede 162 seja chamado quando a lógica de controle 224 determinar que o comando está associado ao serviço da Rede 162. Por exemplo, uma solicitação 260 pode ser enviada ao serviço da Rede 162 através do componente de rede do núcleo do SO 210. O componente de rede pode formatar e encaminhar a solicitação através da rede 170 (tal como encapsulando a solicitação em um pacote de rede de acordo com um protocolo da rede 170) para o serviço da Rede 162 para realizar o

comando do usuário. A solicitação 260 pode incluir múltiplas etapas, tal como abertura de um canal de comunicação (por exemplo, um soquete) entre a lógica de controle 224 e o serviço de Rede 162, e enviar informações relacionadas ao comando do usuário. O serviço de Rede 162 pode responder à solicitação 260 com uma resposta que pode ser transmitida através da rede 170 e encaminhada pelo componente de rede para a lógica de controle 224 como a resposta 262. A resposta do serviço da Rede 162 pode incluir um estado do serviço da Rede 162 e outras informações em resposta ao comando do usuário.

[0048] A lógica de controle 224 pode gerar uma saída (com o auxílio de um mecanismo de sintetização de saída da IU 228 e o componente de sintetização do núcleo SO 210) para ser apresentada ao usuário baseado nas respostas dos aplicativos. Por exemplo, a estrutura de dados comandada 140 pode mapear estados recebidos a partir das funções para respostas fornecidas ao usuário a partir do assistente digital pessoal controlado por voz 120. Em geral, a lógica de controle 224 pode fornecer comandos de saída de alto nível ao mecanismo de sintetização de saída da IU 228, que pode produzir primitivos de saída de nível inferior ao componente de sintetização do núcleo do SO 210 para saída visual em um meio de exibição, saída de áudio e/ou voz através de um alto-falante ou fone de ouvido, e saída de vibração a partir de um motor elétrico. Por exemplo, a lógica de controle 224 pode enviar um comando de conversão de texto para voz com uma cadeia de caracteres de texto para o mecanismo de sintetização de saída da IU 228, que pode gerar dados de áudio digital simulando uma voz falada.

[0049] A lógica de controle 224 pode determinar quais informações irá fornecer ao usuário baseado em um estado do aplicativo. Os estados podem corresponder ao início, processamento, confirmação, desambiguação ou término de um comando do usuário. A estrutura de dados de comando 140 pode mapear os estados do aplicativo para

diferentes respostas a serem fornecidas aos usuários. Os tipos de informação que podem ser fornecidos incluem texto de exibição, fala simulada, uma ligação profunda de volta ao aplicativo, uma ligação para uma página ou site da Rede, e conteúdo da Rede baseado na Linguagem de Marcação de Hipertexto (HTML), por exemplo.

### **Exemplos de Estados de Aplicativo**

[0050] A Figura 3 é um diagrama de um exemplo de máquina de estado 300 para um aplicativo fazendo interface com o assistente digital pessoal 120 de um jeito sem orientação. O aplicativo pode começar em um estado de pré-carregamento 310 ou em um estado inicial 320. O estado de pré-carregamento 310 pode ser ativado quando o assistente digital pessoal 120 faz com que o aplicativo seja pré-carregado, tal como quando o nome do aplicativo é conhecido, mas o comando falado não está completo. O aplicativo permanecerá no estado de pré-carregamento 310 até que as operações de pré-carregamento estejam completas. Quando as operações de pré-carregamento estiverem completas, o aplicativo pode transitar para o estado inicial 320.

[0051] O estado inicial 320 pode ser ativado após o estado de pré-carregamento 310 ser concluído ou após o comando do usuário ser fornecido pelo assistente digital pessoal 120 para o aplicativo. Durante o estado inicial 320, o comando do usuário é processado pelo aplicativo. Se o comando não for ambíguo, mas levará mais do que uma quantidade de tempo predeterminada para completar (tal como cinco segundos), o estado pode ser mudado para um estado em progresso 330 enquanto o comando está sendo executado. Se o comando não for ambíguo e puder resultar na execução de uma operação importante ou destrutiva, o estado pode ser mudado para um estado de confirmação 340. Se o comando for um tanto ambíguo, mas a ambiguidade puder ser esclarecida pela escolha entre algumas opções, o estado pode ser mudado para um estado de desambiguação 350. Se o comando for

ambíguo e não puder ser desambiguado com algumas opções, o estado pode ser mudado para um estado final 360, tal como um estado de falha ou um estado de redirecionamento. Se o comando não puder ser realizado, o estado pode ser mudado para um estado final 360, tal como o estado de falha. Se o comando puder ser concluído em menos do que uma quantidade de tempo predeterminada para conclusão e não for desejado solicitar confirmação do usuário, o estado pode ser mudado para um estado final 360, tal como um estado de sucesso. Deve-se observar que o estado final 360 pode ser um único estado com múltiplas condições (tal como em que as condições são sucesso, falha, redirecionamento e tempo limite), ou um grupo de estados finais (tal como em que os estados são sucesso, falha, redirecionamento e tempo limite).

[0052] O estado de progresso 330 pode indicar que as operações do comando do usuário estão sendo realizadas ou estão sendo tentadas. O aplicativo pode fornecer informações ao usuário durante o estado de progresso 330 por enviar uma cadeia de caracteres de conversão de texto para fala (TTS) ou uma cadeia de caracteres da interface gráfica do usuário (GUI) para o assistente digital pessoal 120 de modo que as informações possam ser apresentadas ao usuário usando a interface do usuário do assistente digital pessoal 120. Adicionalmente ou como alternativa, informações-padrão (tal como uma roda giratória, uma ampolheta e/ou um botão cancelar) podem ser apresentadas ao usuário durante o estado de progresso 330 usando a interface do usuário do assistente digital pessoal 120.

[0053] Durante o estado de progresso 330, o aplicativo pode monitorar o progresso das operações e determinar se o aplicativo pode permanecer no estado de progresso 330 ou transitar para o estado final 360. Em uma concretização, o aplicativo pode iniciar um temporizador (tal como por cinco segundos) e, se o aplicativo não tiver progredido o

suficiente antes de o temporizador expirar, o estado pode ser mudado para o estado final 360, tal como um estado de tempo limite. Se o aplicativo estiver obtendo progresso suficiente, o temporizador pode ser reiniciado e o progresso pode ser novamente examinado na próxima vez que o temporizador expirar. O aplicativo pode ter um limite de tempo máximo para permanecer no estado de progresso 330, e se o limite de tempo máximo for excedido, o estado pode ser mudado para o estado final 360, tal como o estado de tempo limite. As operações associadas ao comando do usuário podem ser concluídas (seja com êxito ou não) e o estado pode ser mudado para o estado final apropriado 360. O usuário pode terminar o aplicativo quando ele está no estado de progresso 330 fornecendo um comando à interface do usuário do assistente digital pessoal 120. Por exemplo, o usuário pode pressionar ou clicar em um botão "cancelar" ou "retornar" em um meio de exibição ou dizer "cancelar". O cancelamento do comando pode fazer com que o assistente digital pessoal 120 interrompa o aplicativo, e exiba uma tela inicial do assistente digital pessoal 120 ou saia.

[0054] O estado de confirmação 340 pode indicar que o aplicativo está aguardando por confirmação do usuário antes de completar uma tarefa. Quando o assistente digital pessoal 120 detecta que o aplicativo está no estado de confirmação 340, uma solicitação por uma resposta sim/não pode ser apresentada ao usuário usando a interface de usuário do assistente digital pessoal 120. O aplicativo pode fornecer ao assistente digital pessoal 120 uma cadeia de caracteres TTS que é uma pergunta com uma resposta de sim ou não. O assistente digital pessoal 120 pode falar a cadeia de caracteres TTS fornecida pelo aplicativo e pode escutar por uma resposta "Sim/Não". Se a resposta do usuário não se resolver em uma resposta sim ou não, o assistente digital pessoal 120 pode continuar a fazer a pergunta para o usuário um número de vezes predefinido (tal como três vezes). Se todas as tentativas tiverem

se esgotado, o assistente digital pessoal 120 pode dizer uma frase-padrão, tal como "Desculpe, não entendi. Toque embaixo para escolher uma resposta", e o assistente digital pessoal 120 pode parar de escutar. Se o usuário tocar em sim ou não, o assistente digital pessoal 120 pode enviar a escolha do usuário ao aplicativo. Se o usuário tocar em um ícone de microfone, o assistente digital pessoal 120 pode tentar reconhecer novamente uma resposta falada (tal como reiniciando um contador que conta o número de tentativas de responder verbalmente). O assistente digital pessoal 120 pode realizar um *loop* até que exista uma correlação ou o usuário cancele ou aperte o botão voltar na tela de exibição. Se o aplicativo receber uma resposta afirmativa do assistente digital pessoal 120, o aplicativo pode tentar completar a tarefa. Se a tarefa for concluída com êxito, o estado pode transitar para o estado final 360 com uma condição de sucesso. Se a tarefa falhar em completar com êxito ou o pedido for cancelado, o estado pode transitar para o estado final 360 com uma condição de falha. Se a tarefa for levar mais do que uma quantidade de tempo predeterminada para ser concluída, o estado pode ser mudado para o estado de progresso 330 enquanto a tarefa está sendo realizada.

[0055] O estado de desambiguação 350 pode indicar que o aplicativo está aguardando o usuário esclarecer entre um número limitado (tal como dez ou menos) de opções antes de completar uma tarefa. O aplicativo pode fornecer ao assistente digital pessoal 120 uma cadeia de caracteres TTS, uma cadeia de caracteres GUI e/ou uma lista de itens da qual o usuário irá escolher. A lista de itens pode ser fornecida como um modelo com uma ou mais unidades de informação para fornecer ao usuário para cada item, tal como um título, uma descrição e/ou um ícone. O assistente digital pessoal 120 pode apresentar a lista de itens ao usuário usando as informações fornecidas pelo aplicativo. O assistente digital pessoal 120 pode solicitar e escutar por uma seleção

do usuário. O usuário pode selecionar a partir da lista usando seleção flexível ou não flexível. A seleção não flexível significa que o usuário somente pode selecionar a partir da lista de uma maneira, ao passo que a seleção flexível significa que o usuário pode selecionar a partir da lista de diversas maneiras diferentes. Por exemplo, o usuário pode selecionar a partir da lista baseado na ordem numérica na qual os itens são listados, tal como falando "primeiro" ou "segundo" para selecionar o primeiro item ou o segundo item, respectivamente. Como outro exemplo, o usuário pode selecionar a partir da lista baseado em relações espaciais entre os itens, tal como "o de baixo", "o de cima", "o da direita" ou "o segundo de baixo para cima". Como outro exemplo, o usuário pode selecionar a partir da lista dizendo o título do item.

[0056] Como um exemplo específico de desambiguação, o usuário pode falar, para o assistente digital pessoal 120, "Aplicativo de Filme, adicione o Filme X à minha fila". Entretanto, pode haver três versões do Filme X, tal como o original e duas continuações: Filme X I, Filme X II e Filme X III. Em resposta ao comando falado, o assistente digital pessoal 120 pode inicializar o Aplicativo de Filme no segundo plano com o comando de adicionar Filme X à fila. O Aplicativo de Filme pode pesquisar pelo Filme X e determinar que há três versões. Assim, o Aplicativo de Filme pode transitar para o estado de desambiguação 350 e enviar as três escolhas alternativas ao assistente digital pessoal 120. O assistente digital pessoal 120, através de sua interface de usuário, pode apresentar três escolhas ao usuário, e o usuário pode selecionar uma a partir da lista. Quando uma seleção apropriada é feita pelo usuário, o assistente pessoal digital 120 pode enviar a resposta ao Aplicativo de Filme e o filme correto pode ser adicionado à fila.

[0057] Se a resposta do usuário não puder ser resolvida em um item na lista, o assistente digital pessoal 120 pode continuar a fazer a pergunta para o usuário até um número de vezes predefinido. Se todas



as tentativas tiverem se esgotado, o assistente digital pessoal 120 pode dizer uma frase padrão, tal como "Desculpe, não entendi. Toque embaixo para escolher uma resposta", e o assistente digital pessoal 120 pode parar de escutar. Se o usuário tocar em um dos itens na lista exibida, o assistente digital pessoal 120 pode enviar a escolha do usuário ao aplicativo. Se o usuário tocar em um ícone de microfone, o assistente digital pessoal 120 pode tentar reconhecer novamente uma resposta falada (tal como reiniciando um contador que conta o número de tentativas de responder verbalmente). O assistente digital pessoal 120 pode realizar um *loop* até que exista uma correlação ou o usuário cancele ou aperte o botão voltar na tela de exibição. Se o aplicativo receber uma resposta válida do assistente digital pessoal 120, o aplicativo pode tentar completar a tarefa. Se a tarefa necessitar de confirmação do usuário antes de ser realizada, o estado pode transitar para o estado de confirmação 340. Se a tarefa for concluída com êxito, o estado pode transitar para o estado final 360 com uma condição de sucesso. Se a tarefa falhar em completar com êxito ou o pedido for cancelado, o estado pode transitar para o estado final 360 com uma condição de falha. Se a tarefa for levar mais do que uma quantidade de tempo predeterminada para ser concluída, o estado pode ser mudado para o estado de progresso 330 enquanto a tarefa está sendo realizada.

[0058] Deve-se compreender que a máquina de estados 300 ilustrativa pode ser estendida com estados adicionais ou alternativos para permitir várias conversas de múltiplos turnos entre o usuário e um aplicativo. A desambiguação (por meio do estado de desambiguação 350) e a confirmação (por meio do estado de confirmação 340) são exemplos específicos de uma conversa de múltiplos turnos. Geralmente, em uma conversa de múltiplos turnos, um aplicativo sem orientação pode solicitar informações adicionais do usuário sem trazer sua interface de usuário à superfície. Em vez disso, as informações

podem ser obtidas do usuário pelo assistente digital pessoal 120 em nome do aplicativo. Assim, o assistente digital pessoal 120 pode atuar como um conduto entre o usuário e o aplicativo.

[0059] O estado final 360 pode indicar que o aplicativo completou a tarefa com êxito, falhou em completar a tarefa, esgotou o tempo limite, ou está sugerindo que o aplicativo deve ser inicializado no primeiro plano (redirecionamento). Como descrito acima, o estado final 360 pode ser um estado único com múltiplas condições (por exemplo, sucesso, falha, redirecionamento e tempo limite esgotado), ou um grupo de estados finais (por exemplo, sucesso, falha, redirecionamento e tempo limite esgotado). O aplicativo pode fornecer ao assistente digital pessoal 120 uma cadeia de caracteres TTS, uma cadeia de caracteres GUI, uma lista de itens (fornecida por meio de um modelo) e/ou um parâmetro de inicialização. O assistente digital pessoal 120 pode apresentar as informações fornecidas pelo aplicativo ao usuário usando a interface de usuário do assistente digital pessoal 120. Adicionalmente ou como alternativa, o assistente pessoal digital 120 pode apresentar respostas predefinidas ou pré-gravadas associadas às diferentes condições. Por exemplo, se um esgotamento do tempo limite ocorrer ou a tarefa falhar, o assistente digital pessoal 120 pode dizer "Desculpe! Não consegui fazer isto para você. Poderia por favor tentar mais uma vez mais tarde?". Como outro exemplo, se o aplicativo estiver solicitando redirecionamento, o assistente digital pessoal 120 pode dizer "Desculpe. <NomeDoAplicativo> não está respondendo. Inicializando <NomeDoAplicativo>", e o assistente digital pessoal 120 pode tentar inicializar o aplicativo no primeiro plano com o comando de voz original e o parâmetro de inicialização (se um parâmetro de inicialização for fornecido pelo aplicativo). Como outro exemplo, se o aplicativo completar a tarefa com êxito, o assistente digital pessoal 120 pode dizer "Fiz o que você pediu".

**Exemplo de Definição de Comando**

[0060] A Figura 4 é um exemplo de uma definição de comando 400 conformando-se com um esquema que pode ser usado para criar uma estrutura de dados, tal como a estrutura de dados de comando 140, para possibilitar uma interface entre um aplicativo de terceiros e o assistente digital pessoal 120. A definição de comando 400 pode ser escrita em diversas linguagens, tal como a Linguagem de Marcação Extensível (XML) ou em um subconjunto do XML que é definido por um esquema. Por exemplo, o esquema pode definir a estrutura da definição de comando, tais como os elementos legais, a hierarquia dos elementos, os atributos legais e opcionais para cada elemento, e outros critérios adequados. A definição de comando 400 pode ser usada pelo assistente digital pessoal 120 para auxiliar na análise de uma expressão vocal do usuário em diferentes componentes, tal como um aplicativo, um comando ou tarefa, e um item de dados ou atributo, em que o item de dados é opcional. Por exemplo, o comando "MovieAppService, add MovieX to my queue" pode ser dividido em um aplicativo ("MovieAppService"), um comando ("Add"), e um item de dados ("MovieX"). A definição de comando 400 pode incluir elementos para definir um nome de aplicativo, tarefas ou comandos do aplicativo, expressões alternativas para processamento de linguagem natural, e respostas associadas a diferentes estados do aplicativo.

[0061] Um ou mais aplicativos podem ser definidos na definição de comando 400. Os aplicativos podem ser aplicativos de terceiros, ou outros, que estão instalados no dispositivo de computação ou nos serviços de Rede. As informações relacionadas ao aplicativo podem ser demarcadas com um elemento definindo o aplicativo. Por exemplo, o nome do aplicativo pode ser definido por um elemento <AppName> e os elementos entre os elementos <AppName> podem ser associados ao elemento principal <AppName>. Na definição de comando 400, o

nome do aplicativo é "MovieAppService", e os elementos que sucedem o elemento `<AppName>` são associados ao aplicativo "MovieAppService".

[0062] Os comandos que seguem o nome do aplicativo são os comandos do aplicativo. Os comandos podem ser identificados com um elemento `<Command>`. Os atributos do elemento *command* podem incluir um nome (por exemplo, "Name") do comando e um tipo de ativação (por exemplo, "ActivationType") do comando. Por exemplo, o tipo de ativação pode ser "foreground" para comandos que deverão ser inicializados no primeiro plano e o tipo de ativação pode ser "segundo plano" para comandos que deverão ser inicializados no segundo plano. O atributo "ActivationType" pode ser opcional, com o tipo de ativação padrão sendo o primeiro plano.

[0063] O elemento `<ListenFor>` pode ser aninhado dentro do elemento `<Command>` e pode ser usado para definir uma ou mais maneiras nas quais o comando pode ser falado. Palavras opcionais ou de suporte podem ser fornecidas como pistas ao assistente digital pessoal 120 durante a execução do processamento de linguagem natural. As palavras de suporte podem ser identificadas dentro de colchetes: [ ]. Os itens de dados podem ser identificados dentro de colchetes: { }. Na definição de comando 400, geralmente há duas formas alternativas de chamar o comando "Add", conforme definido pelos dois elementos `<ListenFor>`. Por exemplo, pode-se dizer tanto "add MovieX to my queue" quanto "add MovieX to my MovieAppService queue" para fazer o assistente digital pessoal 120 inicializar o comando "Add" do MovieAppService no segundo plano. Deve-se observar que expressões predefinidas podem ser identificadas com a palavra-chave "builtIn:" dentro de um par de colchetes : {builtIn:<identificador de expressão>}.

[0064] O elemento `<Feedback>` pode ser aninhado dentro do elemento `<Command>` e pode ser usado para definir uma expressão a

ser falada para o usuário quando o assistente digital pessoal 120 reconheceu com êxito um comando falado pelo usuário. Adicionalmente ou como alternativa, o elemento <Feedback> pode definir uma cadeia de caracteres de texto a ser exibida para o usuário à medida que o comando falado está sendo analisado pelo assistente digital pessoal 120.

[0065] O elemento <Response> pode ser aninhado dentro do elemento <Command> e pode ser usado para definir uma ou mais respostas fornecidas pelo assistente digital pessoal 120 ao usuário. Cada resposta é associada a um estado do aplicativo, conforme definido por um atributo "State". Os estados podem ser para estados finais, tal como sucesso e falha, ou para estados intermediários, tal como progresso. Pode haver múltiplos tipos de respostas definidas, tal como <DisplayString> para exibir texto em uma tela, <TTSString> para o texto que será falado para o usuário, <AppDeepLink> para uma ligação profunda com um site da Rede, e <WebLink> para uma ligação menos profunda para um site da Rede, por exemplo. As respostas definidas pelo elemento <Response> podem ser aumentadas com informações de resposta adicionais fornecidas pelo pedido.

### **Exemplo de Diagrama de Sequência**

[0066] A Figura 5 é um exemplo de diagrama de sequência 500 ilustrando a comunicação de múltiplos encadeamentos de execução (510, 520 e 530) para realizar, sem orientação, uma função de um aplicativo de terceiros de dentro do assistente digital pessoal 120. O encadeamento da IU 510 e o encadeamento de controle 520 podem ser encadeamentos paralelos de uma concretização de múltiplos encadeamentos do assistente digital pessoal 120. O encadeamento da IU 510 pode ser primariamente responsável por capturar a entrada a partir da, e exibir saída para a interface do usuário do assistente digital pessoal 120. Por exemplo, a entrada de fala, a entrada tátil e/ou a

entrada de texto podem ser capturadas pelo encadeamento da IU 510. Em uma concretização, o encadeamento da IU 510 pode realizar o processamento de linguagem natural na entrada e pode correlacionar os comandos falados pelo usuário em comandos na estrutura de dados de comando 140. Quando o comando falado é determinado como correlacionando um comando na estrutura de dados de comando 140, o comando pode ser comunicado ao encadeamento de controle 520 para processamento adicional. Em uma concretização alternativa, o encadeamento da IU 510 pode capturar a fala em entrada de texto, e as palavras individuais podem ser comunicadas ao encadeamento de controle 520, que pode realizar o processamento de linguagem natural na entrada e pode correlacionar os comandos falados pelo usuário aos comandos na estrutura de dados de comando 140.

[0067] O encadeamento de controle 520 pode ser primariamente responsável por se comunicar com o acompanhar o progresso do aplicativo e fazer interface com o encadeamento da IU 510. Por exemplo, o encadeamento de controle 520 pode ser notificado pelo encadeamento da IU 510 de que o usuário falou para a interface de usuário do assistente digital pessoal 120. As palavras ou comandos podem ser recebidos pelo encadeamento de controle 520 e o encadeamento de controle 520 pode notificar o encadeamento da IU 510 quando um comando do usuário tiver sido reconhecido pelo encadeamento de controle 520. O encadeamento da IU 510 pode indicar ao usuário, por meio da interface de usuário do assistente pessoal digital 120, que o progresso está sendo feito no comando. O encadeamento da IU 510 ou o encadeamento de controle 520 podem determinar que o comando deverá ser inicializado sem orientação, recuperando atributos do comando a partir da estrutura de dados de comando 140. O encadeamento de controle 520 pode iniciar um novo encadeamento ou se comunicar com um encadeamento existente, tal

como o encadeamento AppService 530, quando o comando tiver que ser inicializado sem orientação. Para reduzir o tempo de resposta ao usuário, pode ser desejável que o encadeamento AppService 530 seja um encadeamento existente, em vez de fazer o encadeamento de controle 520 iniciar um novo encadeamento. Por exemplo, o encadeamento AppService 530 pode ser inicializado ao aquecer o aplicativo ou durante uma inicialização do dispositivo de computação 130.

[0068] O encadeamento AppService 530 pode ser executado no dispositivo de computação 130 ou pode ser executado em um servidor remoto, tal como o computador servidor remoto 160. O encadeamento AppService 530 pode ser primariamente responsável por completar a função especificada pelo comando do usuário. O encadeamento AppService 530 pode manter uma máquina de estados (tal como a máquina de estados 300) para acompanhar o progresso de execução da função, e pode fornecer atualizações sobre a condição ao encadeamento de controle 520. Por fornecer atualizações de condição ao encadeamento de controle 520, o encadeamento AppService 530 pode dispensar o uso de orientação, sendo que a saída para o usuário é fornecida pelo assistente digital pessoal 120 e não por uma interface de usuário do encadeamento AppService 530.

[0069] O encadeamento de controle 520 pode acompanhar o progresso do aplicativo (por exemplo, o encadeamento App Service 530) por receber atualizações de condição a partir do aplicativo e verificar se o aplicativo está obtendo progresso. Por exemplo, o encadeamento de controle 520 pode iniciar um temporizador com uma duração predefinida (tal como cinco segundos) toda vez que ele se comunicar com o encadeamento AppService 530 (seja enviando informações ao encadeamento AppService 530 ou recebendo informações do encadeamento AppService 530). Se o temporizador

expirar antes de o encadeamento AppService 530 responder, o encadeamento de controle 520 pode indicar ao encadeamento da IU 510 que o aplicativo falhou em responder e o encadeamento da IU 510 pode apresentar uma mensagem de falha ao usuário por meio da interface de usuário do assistente digital pessoal 120. O encadeamento AppService 530 pode ser terminado ou ignorado pelo encadeamento de controle 520 após o temporizador expirar. Como alternativa, se o encadeamento AppService 530 responder antes de o temporizador expirar, o temporizador pode ser reiniciado se outra resposta for esperada do aplicativo (tal como quando o aplicativo responde com o estado do progresso), ou o temporizador pode ser cancelado (tal como quando o aplicativo completou a função (um estado final), ou quando uma resposta do usuário está sendo solicitada (um estado de confirmação ou desambiguação)).

[0070] Quando o encadeamento de controle 520 recebe um estado de confirmação ou desambiguação a partir do encadeamento AppService 530, o encadeamento de controle 520 pode indicar, ao encadeamento da IU 510, que a confirmação ou desambiguação é solicitada pelo usuário. O encadeamento da IU 510 pode apresentar as escolhas de confirmação ou desambiguação ao usuário por meio da interface de usuário do assistente digital pessoal 120. Quando o usuário responder, ou falhar em responder, o encadeamento da IU 510 pode fornecer ao usuário resposta, ou a falta definitiva da mesma, ao encadeamento de controle 520. O encadeamento de controle 520 pode passar a resposta do usuário ao encadeamento AppService 530 de modo que o encadeamento AppService 530 possa realizar a função. Se o usuário falhar em responder, o encadeamento de controle 520 pode terminar o encadeamento AppService 530.

[0071] O encadeamento da IU 510 pode exibir vários tipos de saída por meio da interface de usuário do assistente digital pessoal 120. Por



exemplo, o encadeamento da IU 510 pode gerar saída de áudio, tal como saída de fala digital simulada gerada a partir de texto. A fala digital simulada pode ser enviada a um chip de processamento de áudio que pode converter a fala digital simulada em um sinal analógico (tal como com um conversor digital para analógico) que pode ser emitido como som por meio de um alto-falante ou fone de ouvido. Como outro exemplo, o encadeamento da IU 510 pode fornecer saída visual, tais como imagens, animação, saída de texto e hiperligações para visualização pelo usuário em uma tela de exibição do dispositivo de computação 130. Caso o usuário toque ou clique nas hiperligações, o encadeamento da IU 510 pode iniciar um aplicativo de navegador para visualizar um site da Rede correspondendo à hiperligação selecionada. Como outro exemplo, o encadeamento da IU 510 pode gerar saída tátil, tal como pelo envio de um sinal de vibração para um motor elétrico que pode fazer o dispositivo de computação 130 vibrar.

### **Exemplo de Método para Conclusão de Tarefas sem Orientação**

[0072] A Figura 6 é um fluxograma de um exemplo de método 600 para completar uma tarefa de um aplicativo, sem orientação, no segundo plano do assistente digital pessoal 120. Em 610, uma entrada de voz, gerada por um usuário, pode ser recebida pelo assistente digital pessoal 120. A entrada de voz pode ser capturada localmente no dispositivo de computação 130 ou remotamente a partir do dispositivo de computação 130. Como um exemplo, a entrada de voz gerada pelo usuário pode ser capturada localmente por um microfone 150 do dispositivo de computação 130 e digitalizada por um conversor de analógico para digital. Como outro exemplo, a entrada de voz gerada pelo usuário pode ser capturada remotamente por um microfone (tal como por um dispositivo Bluetooth acompanhante) conectado por tecnologia sem fio ao dispositivo de computação 130. O assistente digital pessoal 120 pode ser controlado pela voz e/ou texto inserido na

interface do usuário do assistente digital pessoal 120.

[0073] Em 620, o processamento de linguagem natural de entrada de voz pode ser realizado para determinar um comando de voz do usuário. O comando de voz do usuário pode incluir uma solicitação para realizar uma função predefinida de um aplicativo, tal como um aplicativo de terceiros com suporte a voz. A função predefinida pode ser identificada usando uma estrutura de dados que define aplicativos e funções de aplicativos que são suportadas pelo assistente digital pessoal 120. Por exemplo, os aplicativos compatíveis podem ser identificados em um arquivo de definição de comando, tal como a definição de comando 400. Por utilizar um arquivo de definição de comando extensível para definir funções dos aplicativos de terceiros que podem ser executadas sem orientação pelo assistente digital pessoal 120, o assistente digital pessoal 120 pode permitir que o usuário realize mais tarefas com a interface de usuário do assistente digital pessoal 120.

[0074] Em 630, o assistente digital pessoal 120 pode fazer o aplicativo executar, sem orientação, a função predefinida sem que uma interface de usuário do aplicativo apareça em um meio de exibição do dispositivo de computação 130. O assistente digital pessoal 120 pode determinar a execução do aplicativo sem orientação uma vez que o aplicativo é definido como sem orientação na estrutura de dados de comando 410, ou uma vez que o usuário está usando o dispositivo de computação em um modo de mãos livres e executar o aplicativo no primeiro plano poderia possivelmente distrair o usuário. Por exemplo, o assistente digital pessoal 120 pode chamar um serviço da Rede para executar a função predefinida do aplicativo. Como outro exemplo, o assistente digital pessoal 120 pode iniciar um novo encadeamento no dispositivo de computação 130 para executar a função predefinida do aplicativo após o comando do usuário ser determinado. Como outro

exemplo, o assistente digital pessoal 120 pode se comunicar com um encadeamento existente, tal como um encadeamento iniciado durante um pré-carregamento do aplicativo, para executar a função predefinida do aplicativo. A função predefinida pode ser executada como um processo de segundo plano. O aplicativo pode monitorar o progresso da função predefinida, tal como rastreando um estado da função predefinida.

[0075] Em 640, uma resposta pode ser recebida a partir do aplicativo indicando um estado associado à função predefinida. Por exemplo, os estados podem incluir estados de carregamento, inicial, progresso, confirmação, desambiguação e final. A resposta pode incluir informações adicionais, tal como uma lista de modelos, uma cadeia de caracteres de texto, uma cadeia de caracteres de texto para fala, uma imagem, uma hiperligação ou outra informação adequada que possa ser exibida para o usuário por meio da interface de usuário do assistente digital pessoal 120.

[0076] Em 650, a interface de usuário do assistente digital pessoal 120 pode fornecer uma resposta ao usuário baseado no estado recebido associado à função predefinida. Dessa maneira, a resposta pode vir de dentro de um contexto da interface de usuário do assistente digital pessoal 120 sem trazer a interface de usuário do aplicativo para a superfície. Adicionalmente, as capacidades de confirmação e desambiguação do assistente digital pessoal 120 podem ser usadas para confirmar e/ou esclarecer um comando do usuário para o aplicativo.

### **Exemplo de Método para Determinar se um Aplicativo deverá ser Pré-Carregado**

[0077] A Figura 7 é um fluxograma de um método ilustrativo 700 para determinar se um aplicativo deverá ser pré-carregado enquanto um usuário está falando para o assistente digital pessoal 120. Em 710, o

usuário pode digitar, expressar ou falar para o assistente digital pessoal 120. O texto ou fala do usuário pode ser analisado usando técnicas de processamento de linguagem natural, e as palavras individuais podem ser reconhecidas a partir da fala. As palavras individuais podem ser analisadas separadamente e dentro da expressão intermediária onde elas são faladas. Por exemplo, o usuário pode dizer "Ei Assistente, MyApp, faça . . ." A palavra "Ei" pode ser uma palavra de suporte e ser descartada. A palavra "Assistente" pode ser usada para fazer com que o assistente digital pessoal 120 saiba que o usuário está solicitando que ele realize uma ação. A palavra "MyApp" pode ser interpretada como um aplicativo.

[0078] Em 720, as palavras digitadas ou faladas podem ser comparadas com as funções nativas do assistente digital pessoal 120 e as funções fornecidas nas definições de comando extensíveis. Coletivamente, as funções nativas e as funções definidas no arquivo de definição de comando podem ser chamadas de "AppServices conhecidos". As palavras faladas podem ser analisadas e comparadas com os AppServices conhecidos como as palavras que estão sendo recitadas. Em outras palavras, a análise da fala pode ocorrer antes de toda a frase ser falada ou digitada pelo usuário. Se nenhum dos AppServices conhecidos for correlacionado, então em 730, o assistente digital pessoal 120 pode abrir um navegador da Rede para recuperar uma página da Rede do mecanismo de busca com uma cadeia de caracteres de busca correspondendo à expressão falada não reconhecida. O controle do programa pode ser transferido para o navegador da Rede de modo que o usuário possa refinar a pesquisa na Rede e/ou visualizar os resultados. No entanto, se um AppService conhecido for correlacionado, então o método 700 pode continuar em 740.

[0079] Em 740, pode ser determinado se o aplicativo AppService é

uma tarefa de primeiro plano ou de segundo plano. Por exemplo, a definição de comando pode incluir um atributo que define o aplicativo AppService como um aplicativo de primeiro plano ou de segundo plano. Se o aplicativo AppService for uma tarefa de primeiro plano 750, o aplicativo AppService pode ser inicializado no primeiro plano e o controle pode ser transferido para o aplicativo AppService para completar o comando. Se o aplicativo AppService for uma tarefa de segundo plano, então o método 700 pode continuar com as etapas paralelas 760 e 770.

[0080] Em 760, o assistente digital pessoal 120 pode fornecer ao usuário informações com respeito à análise da fala. Especificamente, o assistente digital pessoal 120 pode gerar saída para uma tela em progresso da interface do usuário do assistente digital pessoal 120. A saída pode ser definida em um elemento <Feedback>, aninhado dentro de um elemento <Command>, da definição de comando, por exemplo. A saída pode ser uma cadeia de caracteres de texto e pode ser atualizada continuamente conforme o usuário continua a falar.

[0081] Em 770, o assistente digital pessoal 120 pode pré-carregar o aplicativo AppService sem aguardar o a expressão vocal do usuário terminar. O pré-carregamento do aplicativo AppService pode incluir alocar memória, realizar a busca antecipada (pre-fetching) de instruções, estabelecer uma sessão de comunicação, recuperar informações de um banco de dados, iniciar um novo encadeamento, acionar uma interrupção ou outras operações adequadas específicas ao aplicativo. O aplicativo pode ser pré-carregado com base em uma função especulativa. Por exemplo, as instruções correspondendo à função especulativa podem ser buscadas antecipadamente, mesmo se a função não for conhecida com certeza. Por pré-aquecer o aplicativo antes de o usuário completar o comando falado, o tempo para responder ao usuário pode ser potencialmente diminuído.

[0082] Em 780, o assistente digital pessoal 120 pode continuar a analisar o resultado do reconhecimento de fala parcial até que a expressão vocal esteja completa. O término da expressão vocal pode ser detectado com base no comando sendo analisado e/ou com base em uma pausa feita pelo usuário por uma quantidade de tempo predeterminada. Por exemplo, o término do comando "MovieAppService, adicione MovieX à minha fila" pode ser detectado quando a palavra "fila" for reconhecida. Como outro exemplo, o término do comando, "TextApp, envie uma mensagem para minha esposa visando que chegarei tarde para o jantar" pode ser mais difícil de detectar, pois o comando termina com um item de dado de tamanho desconhecido. Assim, uma pausa pode ser usada para indicar, ao assistente digital pessoal 120, que o comando está completo.

[0083] Em 790, o término do comando falado pode ser detectado e o resultado do reconhecimento da fala final pode ser passado para o aplicativo. O aplicativo e o assistente digital pessoal 120 podem se comunicar um com o outro para completar o comando falado, como descrito com referência às Figuras anteriores.

### **Sistemas de Computação**

[0084] A Figura 8 representa um exemplo generalizado de um sistema de computação adequado 800 no qual as inovações descritas podem ser implementadas. O ambiente de computação 800 não tem a intenção de implicar qualquer limitação quanto ao escopo de uso ou funcionalidade, uma vez que as inovações podem ser implementadas em diversos sistemas de computação de uso geral ou especial.

[0085] Com referência à Figura 8, o sistema de computação 800 inclui uma ou mais unidades de processamento 810, 815 e a memória 820, 825. Na Figura 8, esta configuração básica 830 está incluída dentro de uma linha tracejada. As unidades de processamento 810, 815 executam instruções executáveis por computador. Uma unidade de

processamento pode ser uma unidade de processamento central (CPU) de propósito geral, um processador num circuito integrado de aplicação específica (ASIC) ou qualquer outro tipo de processador. Em um sistema multiprocessamento, várias unidades de processamento executam instruções executáveis por computador para aumentar a capacidade de processamento. Por exemplo, a Figura 8 mostra uma unidade de processamento central 810, assim como uma unidade de processamento de gráficos ou unidade de coprocessamento 815. A memória tangível 820, 825 pode ser memória volátil (por exemplo, registros, memória cache, RAM), memória não volátil (por exemplo, ROM, EEPROM, memória flash, etc.), ou alguma combinação das duas, acessível pela (s) unidade(s) de processamento. A memória 820, 825 armazena o software 880 implementando uma ou mais inovações aqui descritas, na forma de instruções executáveis por computador adequadas para execução pela(s) unidade(s) de processamento.

[0086] Um sistema de computação pode ter recursos adicionais. Por exemplo, o ambiente de computação 800 inclui o armazenamento 840, um ou mais dispositivos de entrada 850, um ou mais dispositivos de saída 860 e uma ou mais conexões de comunicação 870. Um mecanismo de interconexão (não ilustrado), tal como um barramento, controlador, ou rede, interconecta os componentes do sistema de computação 800. Normalmente, o software do sistema operacional (não ilustrado) oferece um ambiente operacional para outros softwares em execução no sistema de computação 800, e coordena as atividades dos componentes do sistema de computação 800.

[0087] O armazenamento tangível 840 pode ser removível ou não removível, e inclui discos magnéticos, fitas ou cassetes magnéticos, CD-ROMs, CD-RWs, DVDs ou qualquer outro meio que possa ser usado para armazenar informações e que possa ser acessado dentro do sistema de computação 800. O armazenamento 840 armazena

instruções para o software 880 implementando uma ou mais inovações aqui descritas.

[0088] O(s) dispositivo(s) de entrada 850 pode(m) ser um dispositivo de entrada a toque, tal como um teclado, mouse, caneta ou trackball, um dispositivo de entrada de voz, um dispositivo de digitalização, ou outro dispositivo que alimenta informações de entrada para o sistema de computação 800. Para codificação de vídeo, o(s) dispositivo(s) de entrada 850 pode(m) ser uma câmera, placa de vídeo, placa sintonizadora de TV ou dispositivo similar que aceite entrada ou vídeo na forma analógica ou digital, ou um CD-ROM ou CD-RW que lê amostras de vídeo para o sistema de computação 800. O(s) dispositivo(s) de saída 860 pode(m) ser um meio de exibição, impressora, alto-falante, gravador de CD, ou outro dispositivo que proporcione uma saída para o sistema de computação 800.

[0089] As conexões de comunicação 870 permitem comunicação por um meio de comunicação para outra entidade de computação. O meio de comunicação transmite informações, como instruções executáveis por computador, entrada ou saída de áudio ou vídeo, ou outros dados em um sinal de dados modulado. Um sinal de dados modulado é um sinal que tem uma ou mais de suas características ajustadas ou alteradas de tal maneira a codificar as informações no sinal. A título de exemplo, e não de limitação, os meios de comunicação podem utilizar um suporte elétrico, óptico, RF ou de outro tipo.

[0090] As inovações podem ser descritas no contexto geral de instruções executáveis por computador, como as incluídas nos módulos de programa, sendo executadas em um sistema de computação em um processador real ou virtual de destino. Geralmente, os módulos de programa incluem rotinas, programas, bibliotecas, objetos, classes, componentes, estruturas de dados etc., que efetuam tarefas específicas ou implementam tipos de dados abstratos específicos. A funcionalidade



dos módulos de programa pode ser combinada ou dividida entre os módulos de programa conforme desejado nas várias concretizações. As instruções executáveis por computador para módulos de programa podem ser executadas dentro de um sistema de computação local ou distribuído.

[0091] Os termos "sistema" e "dispositivo" são utilizados aqui de maneira intercambiável. A menos que o contexto indique claramente o contrário, nenhum termo implica qualquer limitação de um tipo de sistema de computação ou dispositivo de computação. Em geral, um sistema de computação ou dispositivo de computação pode ser local ou distribuído, e pode incluir qualquer combinação de hardware de propósito especial e/ou hardware de propósito geral com software implementando a funcionalidade aqui descrita.

[0092] Para fins de apresentação, a descrição detalhada usa termos como "determinar" e "usar" para descrever operações de computador em um sistema de computação. Esses termos são abstrações de nível superior para as operações realizadas por um computador, e não devem ser confundidos com os atos realizados por um ser humano. As operações reais do computador que correspondem a esses termos variam, dependendo da implementação.

### **Dispositivo Móvel**

[0093] A Figura 9 é um diagrama de sistema que representa um exemplo de dispositivo móvel 900 incluindo uma variedade de componentes opcionais de hardware e software, ilustrados em geral em 902. Quaisquer componentes 902 no dispositivo móvel podem se comunicar com qualquer outro componente, embora nem todas as conexões sejam ilustradas, para facilidade de ilustração. O dispositivo móvel pode ser qualquer um de uma variedade de dispositivos de computação (por exemplo, telefone celular, smartphone, computador portátil, Assistente Pessoal Digital (PDA), etc.) e pode permitir

comunicações bidirecionais sem fios com uma ou mais redes de comunicações móveis 904, tais como uma rede celular, de satélite ou de outro tipo.

[0094] O dispositivo móvel 900 ilustrado pode incluir um controlador ou processador 910 (por exemplo, processador de sinal, microprocessador, ASIC, ou outros circuitos lógicos de controle e processamento) para executar tarefas tais como codificação de sinal, processamento de dados, processamento de entrada/saída, controle de potência e/ou outras funções. Um sistema operacional 912 pode controlar a alocação e utilização dos componentes 902 e o suporte para o assistente pessoal digital 120 e um ou mais programas de aplicação 914. Os programas de aplicação podem incluir aplicativos de computação móvel comuns (por exemplo, aplicativos de correio eletrônico, calendários, gerenciadores de contatos, navegadores da Rede, aplicativos de mensagens, aplicativos de filmes, aplicativos bancários) ou qualquer outro aplicativos de computação. Os programas de aplicação 914 podem incluir aplicativos possuindo tarefas que podem ser executadas sem orientação pelo assistente pessoal digital 120. Por exemplo, as tarefas podem ser definidas na estrutura de dados de comando 140. A funcionalidade 913 para acessar uma loja de aplicativos também pode ser utilizada para adquirir e atualizar programas de aplicação 914.

[0095] O dispositivo móvel 900 ilustrado pode incluir a memória 920. A memória 920 pode incluir memória não removível 922 e/ou memória removível 924. A memória não removível 922 pode incluir RAM, ROM, memória flash, um disco rígido, ou outras tecnologias de armazenamento de memória bem conhecidas. A memória removível 924 pode incluir memória flash ou um cartão de Módulo de Identidade de Assinante (SIM), que é bem conhecido em sistemas de comunicação GSM, ou outras tecnologias de armazenamento de memória bem

conhecidas, tais como "cartões inteligentes". A memória 920 pode ser utilizada para armazenar dados e/ou código para executar o sistema operacional 912 e os aplicativos 914. Exemplos de dados podem incluir páginas da Rede, texto, imagens, arquivos de som, dados de vídeo ou outros conjuntos de dados a serem enviados para e/ou recebidos de um ou mais servidores de rede ou outros dispositivos através de uma ou mais redes com ou sem fios. A memória 920 pode ser utilizada para armazenar um identificador de assinante, tal como uma Identidade de Assinante Móvel Internacional (IMSI), e um identificador de equipamento, tal como um Identificador de Equipamento Móvel Internacional (IMEI). Tais identificadores podem ser transmitidos a um servidor de rede para identificar usuários e equipamento.

[0096] O dispositivo móvel 900 pode suportar um ou mais dispositivos de entrada 930, tais como uma tela de toque 932, o microfone 934, a câmera 936, o teclado físico 938 e/ou o trackball 940 e um ou mais dispositivos de saída 950, tal como um alto-falante 952 e um meio de exibição 954. Outros dispositivos de saída possíveis (não ilustrados) podem incluir dispositivos piezelétricos ou outros dispositivos de saída hápticos. Alguns dispositivos podem servir mais de uma função de entrada/saída. Por exemplo, a tela de toque 932 e o meio de exibição 954 podem ser combinados em um único dispositivo de entrada/saída.

[0097] Os dispositivos de entrada 930 podem incluir uma Interface Natural do Usuário (NUI). Uma NUI é qualquer tecnologia de interface que possibilite ao usuário interagir com um dispositivo de forma "natural", livre de restrições artificiais impostas pelos dispositivos de entrada como o mouse, teclado, controle remoto, entre outros. Exemplos de métodos NUI incluem aqueles baseados no reconhecimento de fala, reconhecimento de toque e estilo, reconhecimento de gestos tanto na tela como adjacente à tela, gestos no ar, acompanhamento visual e de cabeça, voz e fala, visão, toque,

gestos e inteligência de máquina. Outros exemplos de uma NUI incluem detecção de gestos de movimento utilizando acelerômetros/giroscópios, reconhecimento facial, telas 3D, acompanhamento visual, de cabeça e olhar, sistemas de realidade virtual e realidade aumentada imersivos, todos os quais proporcionam uma interface mais natural, bem como tecnologias para detectar atividade cerebral utilizando eletrodos de detecção de campo elétrico (EEG e métodos relacionados). Assim, em um exemplo específico, o sistema operacional 912 ou os aplicativos 914 podem compreender software de reconhecimento de fala como parte de uma interface de voz do usuário que permite ao usuário operar o dispositivo 900 por meio de comandos de voz. Além disso, o dispositivo 900 pode compreender dispositivos de entrada e software que permitem interação do usuário por meio de gestos espaciais do usuário, tal como pela detecção e interpretação de gestos para fornecer entrada a um aplicativo de jogo.

[0098] Um modem sem fio 960 pode ser acoplado a uma antena (não ilustrada) e pode oferecer suporte a uma comunicação bidirecional entre o processador 910 e dispositivos externos, como é compreendido na técnica. O modem 960 é ilustrado genericamente e pode incluir um modem celular para comunicação com a rede de comunicação móvel 904 e/ou outros modems baseados em rádio (por exemplo, Bluetooth 964 ou Wi-Fi 962). O modem sem fio 960 é tipicamente configurado para comunicação com uma ou mais redes celulares, tal como uma rede GSM para comunicações de dados e voz dentro de uma única rede celular, entre redes celulares, ou entre o dispositivo móvel e uma rede comutada de telefonia pública (PSTN).

[0099] O dispositivo móvel pode adicionalmente incluir pelo menos uma porta de entrada/saída 980, uma fonte de alimentação 982, um receptor de sistema de navegação por satélite 984, tal como um receptor do Sistema Global de Posicionamento (GPS), um acelerômetro

986 e/ou um conector físico 990, que pode ser uma porta USB, uma porta IEEE 1394 (FireWire) e/ou uma porta RS-232. Os componentes ilustrados 902 não são necessários ou estão todos inclusos, assim como quaisquer componentes podem ser removidos e outros componentes podem ser acrescentados.

### **Ambiente com Suporte à Nuvem**

[00100] A Figura 10 ilustra um exemplo generalizado de um ambiente com suporte à nuvem adequado 1000 no qual as concretizações, técnicas e tecnologias descritas podem ser implementadas. No ambiente ilustrativo 1000, vários tipos de serviços (por exemplo, serviços de computação) são oferecidos por uma nuvem 1010. Por exemplo, a nuvem 1010 pode compreender um conjunto de dispositivos de computação, que podem estar localizados centralmente ou distribuídos, os quais oferecem serviços baseados em nuvem a vários tipos de usuários e dispositivos conectados por meio de uma rede, tal como a Internet. O ambiente de implementação 1000 pode ser usado de diferentes maneiras para realizar tarefas de computação. Por exemplo, algumas tarefas (por exemplo, processamento da entrada do usuário e apresentação de uma interface do usuário) podem ser realizadas em dispositivos de computação locais (por exemplo, nos dispositivos conectados 1030, 1040, 1050), enquanto que outras tarefas (por exemplo, armazenamento de dados a serem usados no processamento subsequente) podem ser realizadas na nuvem 1010.

[00101] No ambiente 1000 ilustrativo, a nuvem 1010 oferece serviços aos dispositivos conectados 1030, 1040, 1050 com uma variedade de capacidades na tela. O dispositivo conectado 1030 representa um dispositivo com uma tela de computador 1035 (por exemplo, uma tela de tamanho médio). Por exemplo, o dispositivo conectado 1030 poderia ser um computador pessoal, tal como um computador de mesa, laptop, notebook, netbook, entre outros. O dispositivo conectado 1040

representa um dispositivo com uma tela de dispositivo móvel 1045 (por exemplo, uma tela de tamanho pequeno). Por exemplo, o dispositivo conectado 1040 poderia ser um telefone móvel, smartphone, assistente pessoal digital, computador tablet, entre outros. O dispositivo conectado 1050 representa um dispositivo com uma tela grande 1055. Por exemplo, o dispositivo conectado 1050 poderia ser uma tela de televisão (por exemplo, uma televisão inteligente) ou outro dispositivo conectado a uma televisão (por exemplo, um decodificador de sinais ou console de jogo), entre outros. Um ou mais dos dispositivos conectados 1030, 1040, 1050 podem incluir capacidades de tela de toque. As telas de toque podem receber entrada de diferentes formas. Por exemplo, as telas de toque capacitivas detectam a entrada de toque quando um objeto (por exemplo, uma ponta do dedo ou caneta Stylus) distorce ou interrompe uma corrente elétrica percorrendo através da superfície. Como outro exemplo, as telas de toque podem utilizar sensores ópticos para detectar a entrada de toque quando os feixes dos sensores ópticos são interrompidos. O contato físico com a superfície da tela não é necessário para que a entrada seja detectada por algumas telas de toque. Os dispositivos sem capacidades de tela também podem ser usados no ambiente 1000 ilustrativo. Por exemplo, a nuvem 1010 pode oferecer serviços para um ou mais computadores (por exemplo, computadores servidores) sem meios de exibição.

[00102] Os serviços podem ser fornecidos pela nuvem 1010 através de provedores de serviços 1020, ou através de outros provedores de serviços online (não ilustrados). Por exemplo, os serviços de nuvem podem ser personalizados para o tamanho de tela, capacidade de exibição e/ou para a capacidade de tela de toque de um determinado dispositivo conectado (por exemplo, os dispositivos conectados 1030, 1040, 1050).

[00103] No ambiente 1000 ilustrativo, a nuvem 1010 fornece as

tecnologias e soluções aqui descritas aos vários dispositivos conectados 1030, 1040, 1050 usando, pelo menos em parte, os provedores de serviços 1020. Por exemplo, os provedores de serviços 1020 podem fornecer uma solução centralizada para vários serviços baseados em nuvem. Os provedores de serviços 1020 podem gerenciar as assinaturas de serviço para os usuários e/ou dispositivos (por exemplo, para os dispositivos conectados 1030, 1040, 1050 e/ou seus respectivos usuários).

### **Implementações Ilustrativas**

[00104] Embora as operações de alguns dos métodos revelados sejam descritas em uma ordem sequencial específica para praticidade de apresentação, deve-se compreender que esta forma de descrição engloba redistribuição, a menos que uma ordenação específica seja exigida por uma linguagem específica apresentada abaixo. Por exemplo, as operações descritas sequencialmente podem, em alguns casos, serem reorganizadas ou realizadas ao mesmo tempo. Além do mais, para fins de simplicidade, as figuras anexas podem não ilustrar as várias formas em que os métodos revelados podem ser usados em conjunto com outros métodos.

[00105] Quaisquer dos métodos revelados podem ser implementados como instruções executáveis por computador ou como um produto de programa de computador armazenado em um ou mais meios de armazenamento legíveis por computador e executados em um dispositivo de computação (por exemplo, qualquer dispositivo de computação disponível, incluindo smartphones ou outros dispositivos móveis que incluem hardware de computação). Os meios legíveis por computador são quaisquer meios tangíveis disponíveis que possam ser acessados dentro de um ambiente de computação (por exemplo, um ou mais discos de mídia óptica, tal como um DVD ou CD, componentes de memória volátil (tal como DRAM ou SRAM), ou componentes de

memória não volátil (tal como memória flash ou unidades de disco rígido)). A título de exemplo e com referência à Figura 8, os meios de armazenamento legíveis por computador incluem a memória 820 e 825, e o armazenamento 840. A título de exemplo e com referência à Figura 9, os meios de armazenamento legíveis por computador incluem a memória e o armazenamento 920, 922 e 924. O termo meio de armazenamento legível por computador não inclui sinais e ondas portadoras. Além disso, o termo meio de armazenamento legível por computador não inclui conexões de comunicação (por exemplo, 870, 960, 962 e 964).

[00106] Quaisquer das instruções executáveis por computador para implementar as técnicas reveladas, bem como quaisquer dados criados e utilizados durante a implementação das concretizações reveladas, podem ser armazenados em um ou mais meios de armazenamento legíveis por computador. As instruções executáveis por computador podem ser parte, por exemplo, de um aplicativo de software dedicado ou de um aplicativo de software que é acessado ou baixado por meio de um navegador da Rede ou outro aplicativo de software (tal como um aplicativo de computação remoto). Tal software pode ser executado, por exemplo, em um único computador local (por exemplo, qualquer computador comercialmente disponível) ou em um ambiente em rede (por exemplo, por meio da Internet, de uma rede de longa distância, uma rede local, uma rede cliente-servidor (tal como uma rede de computação em nuvem), ou outra tal rede) usando um ou mais computadores de rede.

[00107] Por clareza, apenas certos aspectos selecionados das implementações baseadas em software são descritos. Outros detalhes que são bem conhecidos na técnica são omitidos. Por exemplo, deve-se compreender que a tecnologia revelada não se limita a nenhuma linguagem ou programa de computador específico. Por exemplo, a



tecnologia revelada pode ser implementada por software escrito em C++, Java, Perl, JavaScript, Adobe Flash, ou qualquer outra linguagem de programação adequada. De maneira similar, a tecnologia revelada não se limita a qualquer computador ou tipo de hardware em particular. Certos detalhes dos computadores e hardware apropriado são bem conhecidos e não precisam ser apresentados em detalhes nesta descrição.

[00108] Adicionalmente, quaisquer das concretizações baseadas em software (compreendendo, por exemplo, instruções executáveis por computador para fazer um computador realizar qualquer um dos métodos revelados) podem ser enviadas, baixadas ou acessadas remotamente através de um meio de comunicação adequado. Tal meio de comunicação adequado inclui, por exemplo, a Internet, a Rede Ampla Mundial (Web), uma intranet, aplicações de software, cabo (incluindo cabo de fibra óptica), comunicações magnéticas, comunicações eletromagnéticas (incluindo RF, micro-ondas e comunicações por infravermelho), comunicações eletrônicas ou outros tais meios de comunicação.

[00109] Os métodos, aparelhos e sistemas revelados não deverão ser interpretados como limitantes de forma alguma. Em vez disso, a presente descrição direciona-se a todas as características e aspectos novos e não óbvios das várias concretizações reveladas, separadamente e em diversas combinações e subcombinações uns com os outros. Os métodos, aparelhos e sistemas revelados não se limitam a qualquer aspecto ou característica específica ou combinação dos mesmos, tampouco exigem as concretizações reveladas que qualquer de uma ou mais vantagens específicas esteja presente ou que problemas sejam resolvidos.

[00110] As tecnologias de qualquer exemplo podem ser combinadas com as tecnologias descritas em qualquer um ou mais dos outros

exemplos. Em vista das diversas concretizações possíveis às quais os princípios da tecnologia revelada podem ser aplicados, deve-se reconhecer que as concretizações ilustradas são exemplos da tecnologia revelada e não devem ser interpretados como limitações ao âmbito da tecnologia revelada.

## REIVINDICAÇÕES

1. Dispositivo de computação (130, 900) compreendendo:
  - uma unidade de processamento (810, 815, 910);
  - memória (820, 825, 920);
  - um ou mais microfones (150, 934);
  - o dispositivo de computação (130, 900) configurado com um assistente digital pessoal controlado por fala (120), as operações compreendendo:
    - receber entrada de fala gerada por um usuário através de um ou mais microfones (150, 934);
    - realizar reconhecimento de fala usando a entrada de fala para determinar um comando falado, em que o comando falado compreende uma solicitação para realizar uma tarefa (112) que é uma função pré-determinada de um aplicativo de terceiros (110, 230), e
    - em que a tarefa (112) é identificada usando uma estrutura de dados (140) que define funções suportadas por aplicativos de terceiros disponíveis (110, 230) usando entrada de voz;
    - determinar se a tarefa (112) do aplicativo de terceiros (110, 230) é capaz de ser executada sem orientação;
    - fazer com que o aplicativo de terceiros (110, 230) execute a tarefa (112) sem orientação como um processo de segundo plano sem uma interface de usuário (124) do aplicativo de terceiros habilitado para voz (110, 230) aparecendo em uma tela do dispositivo de computação (130, 900), quando é determinado que a tarefa (112) do aplicativo de terceiros (110, 230) é capaz de ser executada sem orientação;
    - receber uma resposta do aplicativo de terceiros (110, 230) indicando um estado associado à tarefa (112); e
    - fornecer, por uma interface de usuário (124) do assistente pessoal digital controlado por fala (120), uma resposta ao usuário com base no estado recebido associado à tarefa (112) de modo que a

resposta venha de dentro de um contexto do usuário interface (124) do assistente pessoal digital controlado por voz (120) sem trazer à tona a interface de usuário (124) do aplicativo de terceiros (110, 230);

**caracterizado pelo fato de que**

as operações ainda compreendem iniciar uma sequência de aquecimento do aplicativo de terceiros (110, 230) enquanto realiza o reconhecimento de fala e antes da conclusão da determinação do comando falado, em que a sequência de aquecimento inclui a alocação de uma parte da memória (820, 825, 920), busca antecipada de instruções, estabelecer uma sessão de comunicação, recuperar informações de um banco de dados, iniciar um novo encadeamento de execução, ou acionar uma interrupção.

2. Dispositivo de computação (130, 900), de acordo com a reivindicação 1, **caracterizado pelo fato de que** determinar que a tarefa (112) do aplicativo de terceiros (110, 230) é capaz de ser executada sem orientação compreende determinar que a estrutura de dados (140) define a tarefa (112) como uma tarefa em segundo plano, ou determinar que o usuário é usando o dispositivo de computação (130, 900) em um modo mãos livres.

3. Dispositivo de computação (130, 900), de acordo com a reivindicação 1 ou 2, **caracterizado pelo fato de que** a estrutura de dados (140) que define tarefas (112) dos aplicativos de terceiros (110, 230) invocáveis por comando falado é baseada em um XML definição de comando (400) definindo um nome de aplicativo dos aplicativos de terceiros (110, 230), tarefas ou comandos dos aplicativos de terceiros (110, 230), fraseado alternativo para processamento de linguagem natural e respostas associadas a diferentes estados de aplicativo dos referidos aplicativos de terceiros (110, 230).

4. Dispositivo de computação (130, 900), de acordo com uma das reivindicações 1 a 3, **caracterizado pelo fato de que** a sequência

de aquecimento do aplicativo de terceiros (110, 230) compreende um fio de controle (520) iniciando um encadeamento de serviço de aplicativo (530) como um processo em segundo plano para executar a tarefa sem orientação (112) quando é determinado que a tarefa (112) do aplicativo de terceiros (110, 230) é capaz de ser executada sem orientação.

5. Dispositivo de computação (130, 900), de acordo com a reivindicação 4, **caracterizado pelo fato de que** receber a entrada de voz gerada pelo usuário através de um ou mais microfones (150, 934) e fornecer a interface de usuário (124) do assistente pessoal digital controlado por voz (120) é realizada por um encadeamento de interface de usuário (510);

em que a linha de controle (520) ou a linha de interface de usuário (510) realiza o referido reconhecimento de voz e determina se a tarefa (112) do aplicativo de terceiros (110, 230) é capaz de ser executada sem orientação; e

em que o encadeamento de controle (520) é executado em paralelo com o encadeamento de interface de usuário (510).

6. Dispositivo de computação (130, 900), de acordo com a reivindicação 5, **caracterizado pelo fato de que** o encadeamento de controle (520) inicia o encadeamento de serviço de aplicativo (530) em um computador servidor remoto (160).

7. Dispositivo de computação (130, 900), de acordo com qualquer uma das reivindicações 1 a 5, as operações ainda compreendendo:

iniciar um temporizador quando o aplicativo de terceiros (110, 230) for executado como um processo em segundo plano; e

encerrando o processo em segundo plano, se o cronômetro expirar.

8. Dispositivo de computação (130, 900), de acordo com

qualquer uma das reivindicações 1 a 6, **caracterizado pelo fato de que** a resposta do aplicativo de terceiros (110, 230) indica um estado de confirmação e a resposta ao usuário baseada no estado de confirmação solicita que o usuário responda com uma resposta de sim ou não.

9. Método, implementado por um dispositivo de computação (130, 900) compreendendo um microfone, o método compreendendo:

receber, por um assistente digital pessoal controlado por voz (120), uma entrada de voz digital gerada por um usuário, em que a entrada de voz digital é recebida através do microfone;

realizar processamento de linguagem natural usando a entrada de voz digital para determinar um comando de voz de usuário, em que o comando de voz do usuário compreende uma solicitação para realizar uma função predefinida de um aplicativo de terceiros com suporte a voz (110, 230), e em que a função predefinida é identificada usando uma estrutura de dados (140) que define funções suportadas por aplicativos de terceiros com suporte a voz (110, 230) disponíveis utilizando entrada de voz,

determinar se a função pré-determinada do aplicativo habilitado para voz de terceiros (110, 230) é capaz de ser executada sem orientação,

fazer com que o aplicativo habilitado para voz de terceiros (110, 230) execute a função predefinida sem orientação como um processo de segundo plano sem uma interface de usuário (124) do aplicativo habilitado para voz de terceiros (110, 230) aparecendo em um exibição do dispositivo de computação (130, 900) quando é determinado que a função pré-determinada do aplicativo habilitado para voz de terceiros (110, 230) é capaz de ser executada sem orientação;

receber uma resposta do aplicativo habilitado para voz de terceiros (110, 230) indicando um estado associado com a função pré-definida; e fornecer, por uma interface de usuário (124) do assistente

peçoal digital controlado por voz (120), uma resposta ao usuário com base no estado recebido associado à função pré-definida de modo que a resposta venha de dentro de um contexto do interface de usuário (124) do assistente pessoal digital controlado por voz (120) sem trazer à tona a interface de usuário (124) do aplicativo habilitado para voz de terceiros;

**caracterizado pelo fato de que**

iniciar uma sequência de aquecimento do aplicativo habilitado para voz de terceiros durante a execução do reconhecimento de fala e antes da conclusão da determinação do comando de voz do usuário, em que a sequência de aquecimento inclui a alocação de uma parte da memória (820, 825, 920), busca antecipada de instruções, estabelecer uma sessão de comunicação, recuperar informações de um banco de dados, iniciar um novo encadeamento de execução, ou acionar uma interrupção.

10. Método, de acordo com a reivindicação 9, **caracterizado pelo fato de que** a sequência de aquecimento do aplicativo de terceiros (110, 230) compreende um encadeamento de controle (520) iniciando um encadeamento de serviço de aplicativo (530) como um processo de segundo plano para executar a tarefa sem orientação (112) quando é determinado que a tarefa (112) do aplicativo de terceiros (110, 230) é capaz de ser executada sem orientação.

11. Método, de acordo com a reivindicação 9 ou 10, **caracterizado pelo fato de que** a estrutura de dados (140) é extensível através de um esquema de comando utilizável para associar solicitações do usuário a funções predefinidas de um ou mais aplicativos habilitados para voz de terceiros (110, 230).

12. Método, de acordo com uma das reivindicações de 9 a 11, **caracterizado pelo fato de que** a estrutura de dados (140) compreende:

um primeiro mapeamento de comandos de voz para funções suportadas por aplicativos habilitados para voz de terceiros disponíveis (110, 230); e

um segundo mapeamento de estados recebidos das funções para respostas fornecidas ao usuário a partir do assistente pessoal digital controlado por voz (120).

13. Meio de armazenamento legível por computador que armazena um método para completar uma tarefa (112) de um aplicativo habilitado por voz (110, 230) dentro do contexto de um assistente pessoal digital controlado por voz (120), **caracterizado pelo fato de que** as operações compreendem:

receber, pelo assistente pessoal digital controlado por voz (120), uma entrada de voz digital gerada por um usuário, em que a entrada de voz digital é recebida por meio de um microfone;

realizar reconhecimento de voz usando a entrada de voz digital para determinar um comando de voz do usuário, em que o comando de voz do usuário compreende uma solicitação para realizar a tarefa (112) que é uma função pré-determinada do aplicativo habilitado para voz (110, 230), e em que a função pré-definida é identificada usando uma estrutura de dados (140) que mapeia comandos de voz do usuário para tarefas (112) de aplicativos habilitados para voz (110, 230);

determinar se a função pré-determinada do aplicativo habilitado para voz de terceiros (110, 230) é capaz de ser executada sem orientação.

quando é determinado que a função pré-determinada do aplicativo habilitado para voz de terceiros (110, 230) é capaz de ser executado sem orientação, fazendo com que o aplicativo habilitado para voz (110, 230) execute a função pré-determinada sem orientação como um processo de segundo plano sem uma interface de usuário (124) do aplicativo habilitado para voz (110, 230) aparecendo em uma tela do



dispositivo de computação (130, 900);

receber uma resposta do aplicativo habilitado para voz (110, 230) indicando um estado associado com a função predefinida; e

fornecer uma resposta ao usuário com base no estado recebido associado à função predefinida de modo que a resposta venha de dentro do contexto do assistente pessoal digital controlado por voz (120) sem uma interface de usuário (124) do sistema de voz aplicação habilitada (110, 230) superfície;

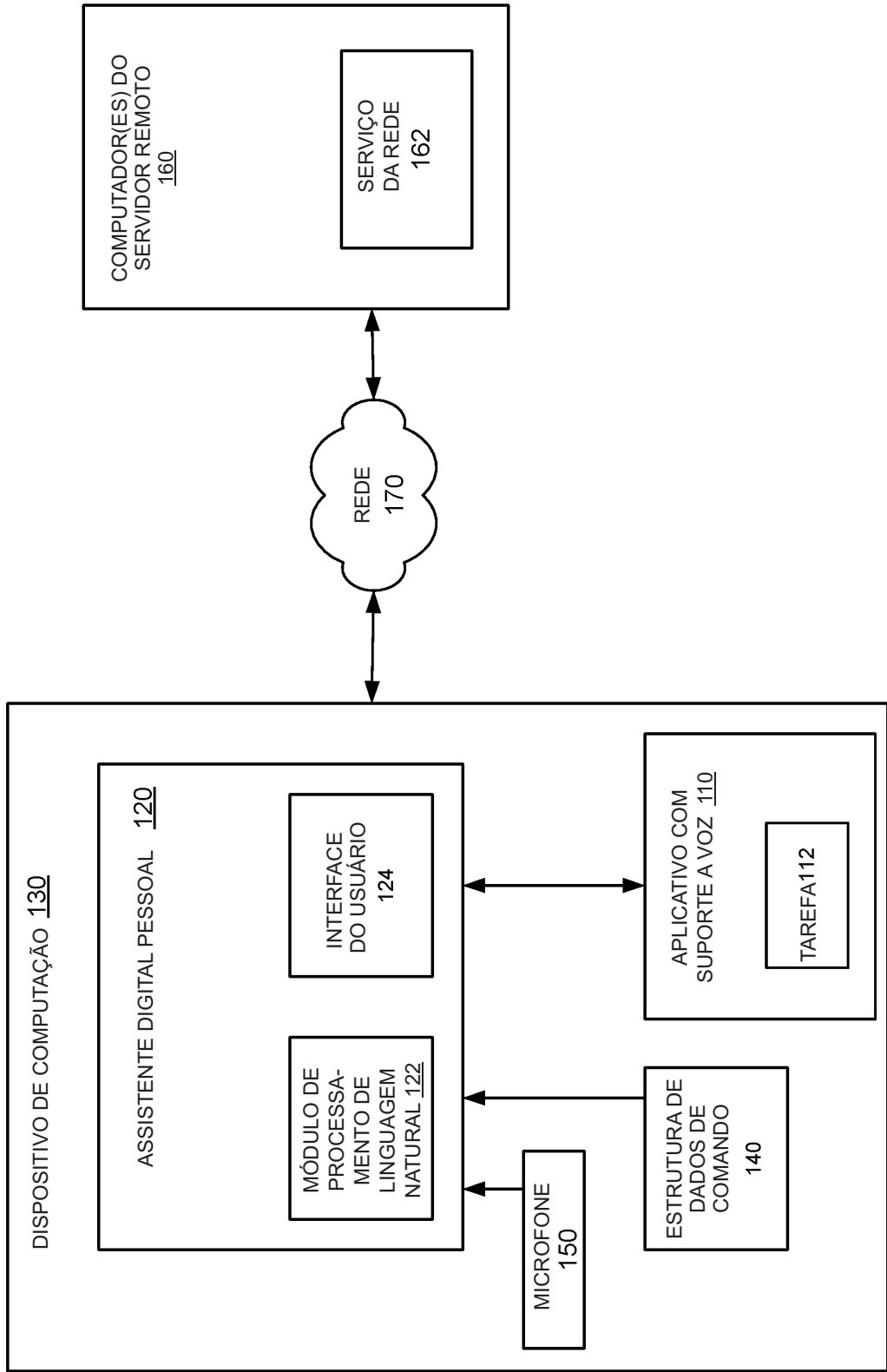
**caracterizado pelo fato de que**

as operações ainda compreendem iniciar uma sequência de aquecimento do aplicativo habilitado para voz (110, 230) enquanto realiza o reconhecimento de voz e antes da conclusão da determinação do comando de voz do usuário, em que a sequência de aquecimento inclui a alocação de uma parte do memória (820, 825, 920), busca antecipada de instruções, estabelecer uma sessão de comunicação, recuperar informações de um banco de dados, iniciar um novo encadeamento de execução, ou acionar uma interrupção.

14. Meio de armazenamento legível por computador, de acordo com a reivindicação 13, **caracterizado pelo fato de que** a sequência de aquecimento do aplicativo de terceiros (110, 230) compreende um encadeamento de controle (520) iniciando um encadeamento de serviço de aplicativo (530) como um processo de segundo plano para executar sem orientação a função predefinida quando for determinado que a função predefinida do aplicativo de terceiros (110, 230) é capaz de ser executada sem orientação.

FIG. 1

100



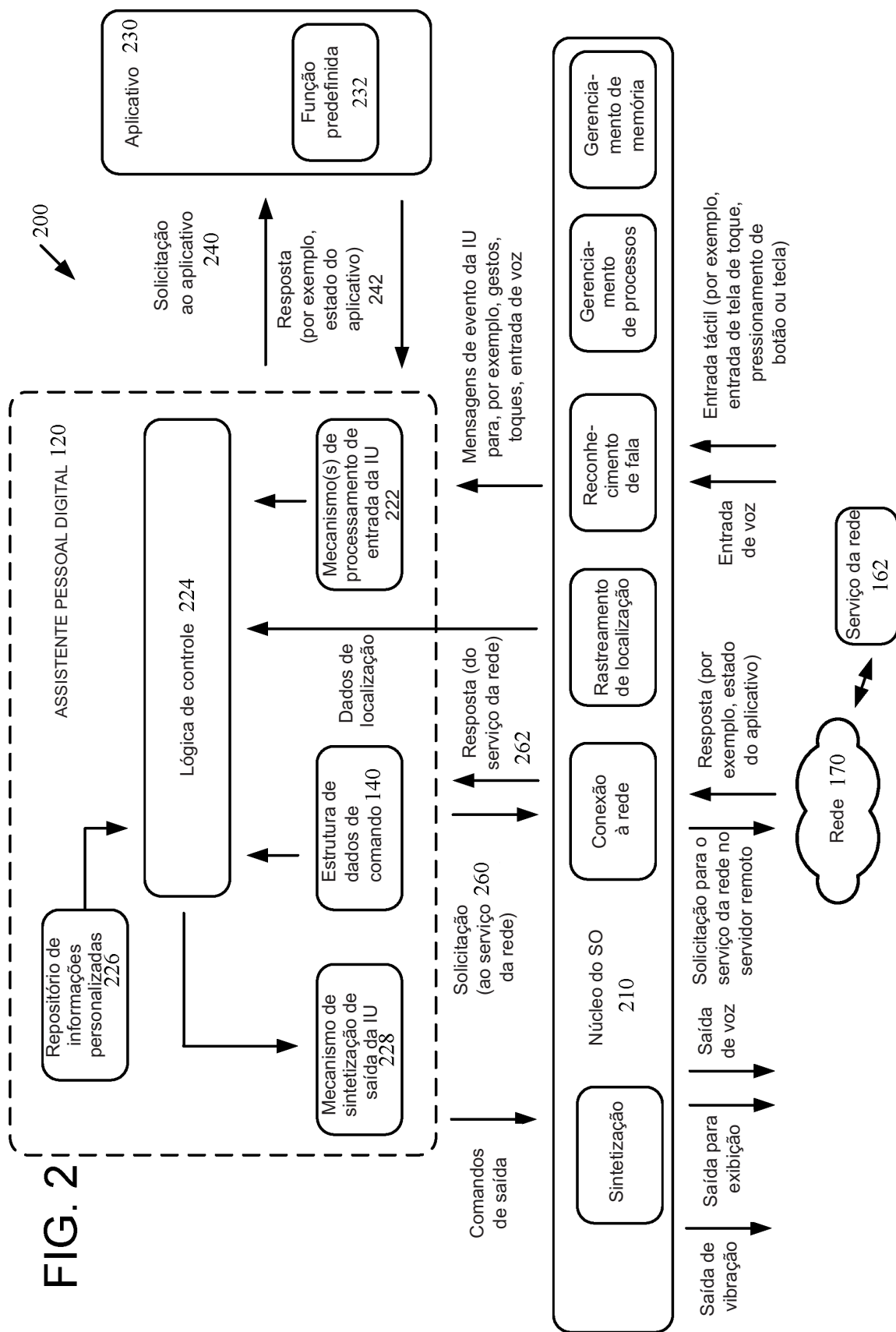
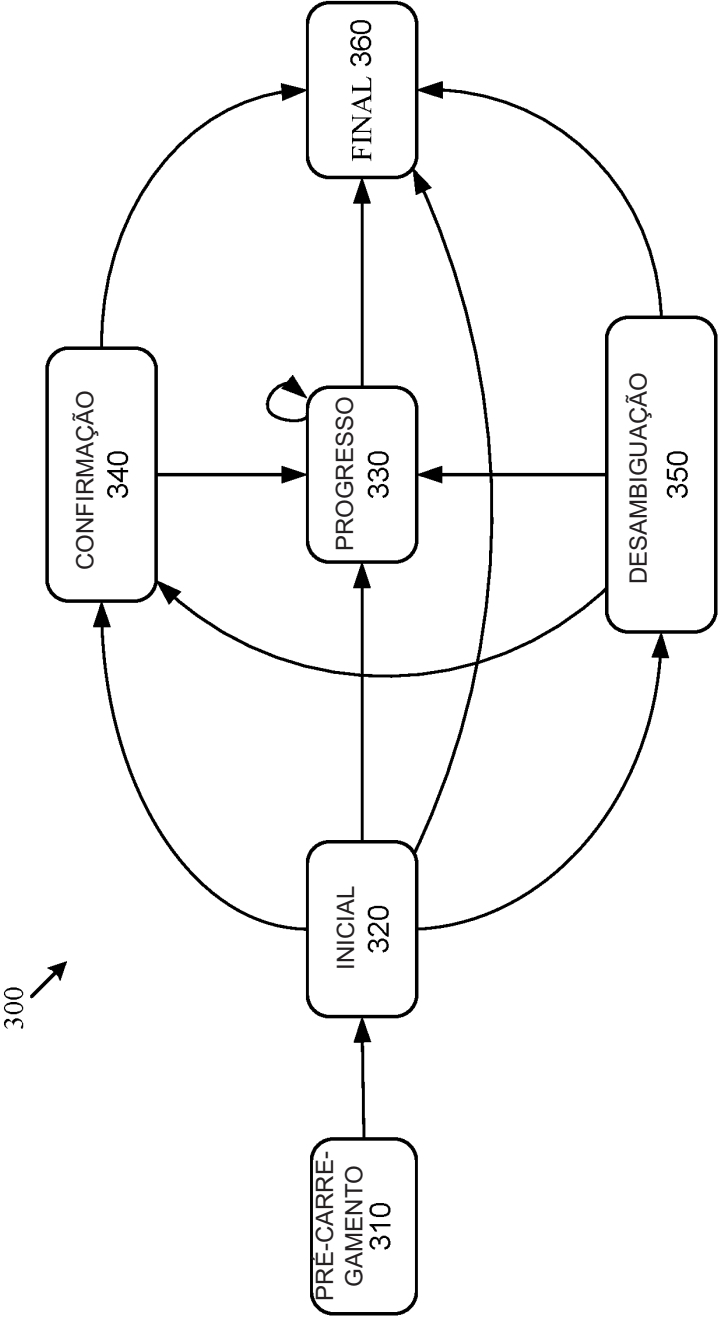



FIG. 3



## FIG. 4

 400

```

<CommandSet Name="MovieAppServiceTest">

  <AppName>MovieAppService</AppName>
  <Example>Add movieName</Example>

  <!-- activationType on a command can be either background or foreground. As an
  example, it can be treated as foreground by default -->
  <Command Name="Add" ActivationType="background">
    <Example>Add the movieName to my MovieAppService queue</Example>
    <ListenFor>Add [the] {movieName} to [my] Queue</ListenFor>
    <ListenFor>Add [the] {movieName} to [my] {builtin:AppName} Queue</ListenFor>
    <Feedback>{builtin:ackPhrase}, adding {movieName} to your queue</Feedback>

    <!-- Responses can be optional and so can individual Response child states. If the
    app does not provide a response for a specific state, a canned response can be displayed.
    -->
    <Responses>

      <Response Name="SuccessMovieAdded" State="Success">
        <DisplayString>{builtin:success}, I've got that done for you</DisplayString>
        <TTSString>{builtin:success}, I've got that done for you</TTSString>
        <AppDeepLink URL="/showQueue.xaml?movieAdded=true">View my MovieAppService
Queue
MovieAppService.com
</WebLink>
        <WebLink URL="http://www.MovieAppService.com/myQueue">Manage queue on
MovieAppService.com
</WebLink>
      </Response>

      <Response Name="FailureMovieNotAdded" State="Failure">
        <DisplayString>{builtin:failure}, that didnt work. Let's try to do this in the
{builtin:appName} app</DisplayString>
        <TTSString>{builtin:failure}, that didnt work. Let's try to do this in the
{builtin:appName} app</TTSString>
        <AppDeepLink></AppDeepLink>
        <WebLink></WebLink>
      </Response>

    </Responses>

    <!-- In case of Headless invocation, the Navigate element can point to the appService to
    launch when this command is hit -->
    <Navigate Target="AppName-VoiceCommandService"/>
  </Command>
</CommandSet>

```

FIG. 5

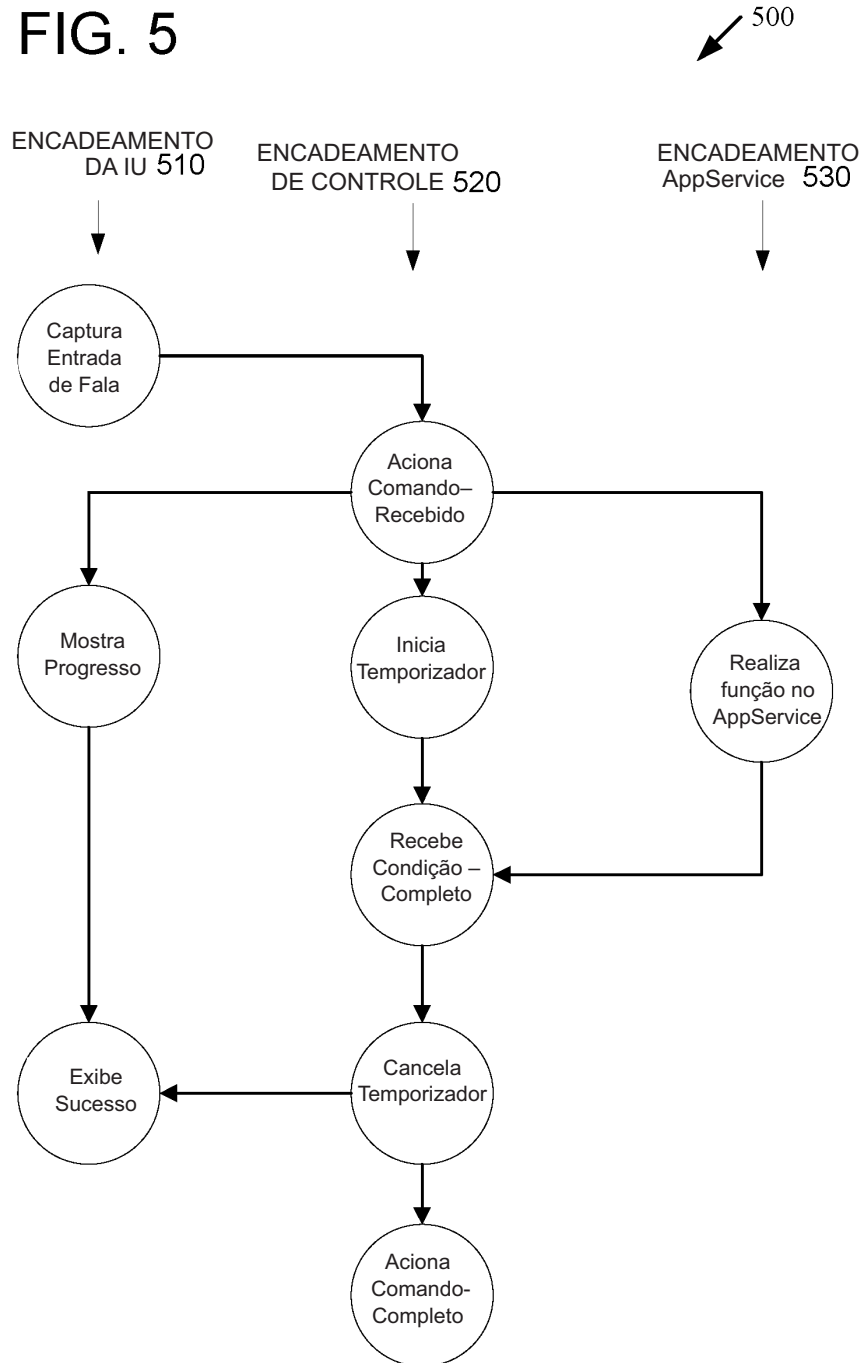


FIG. 6

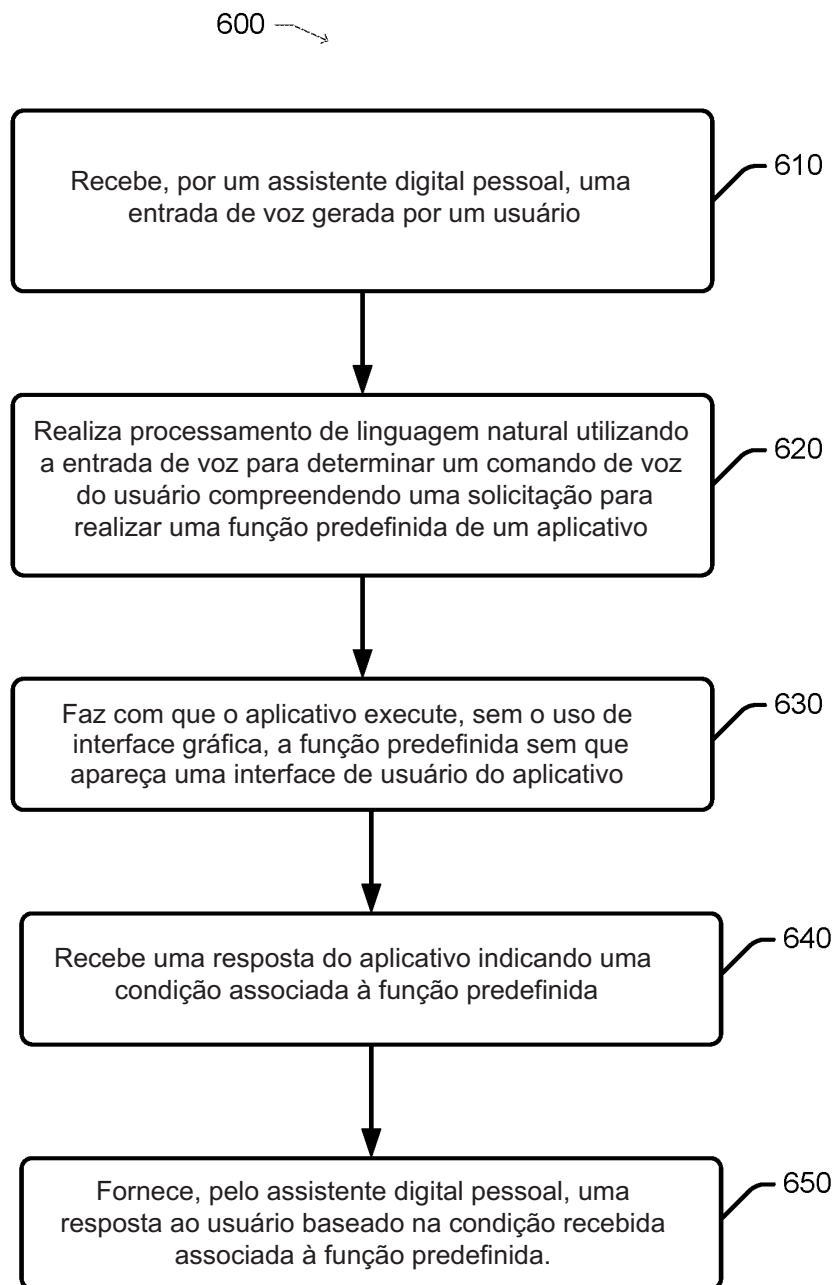


FIG. 7

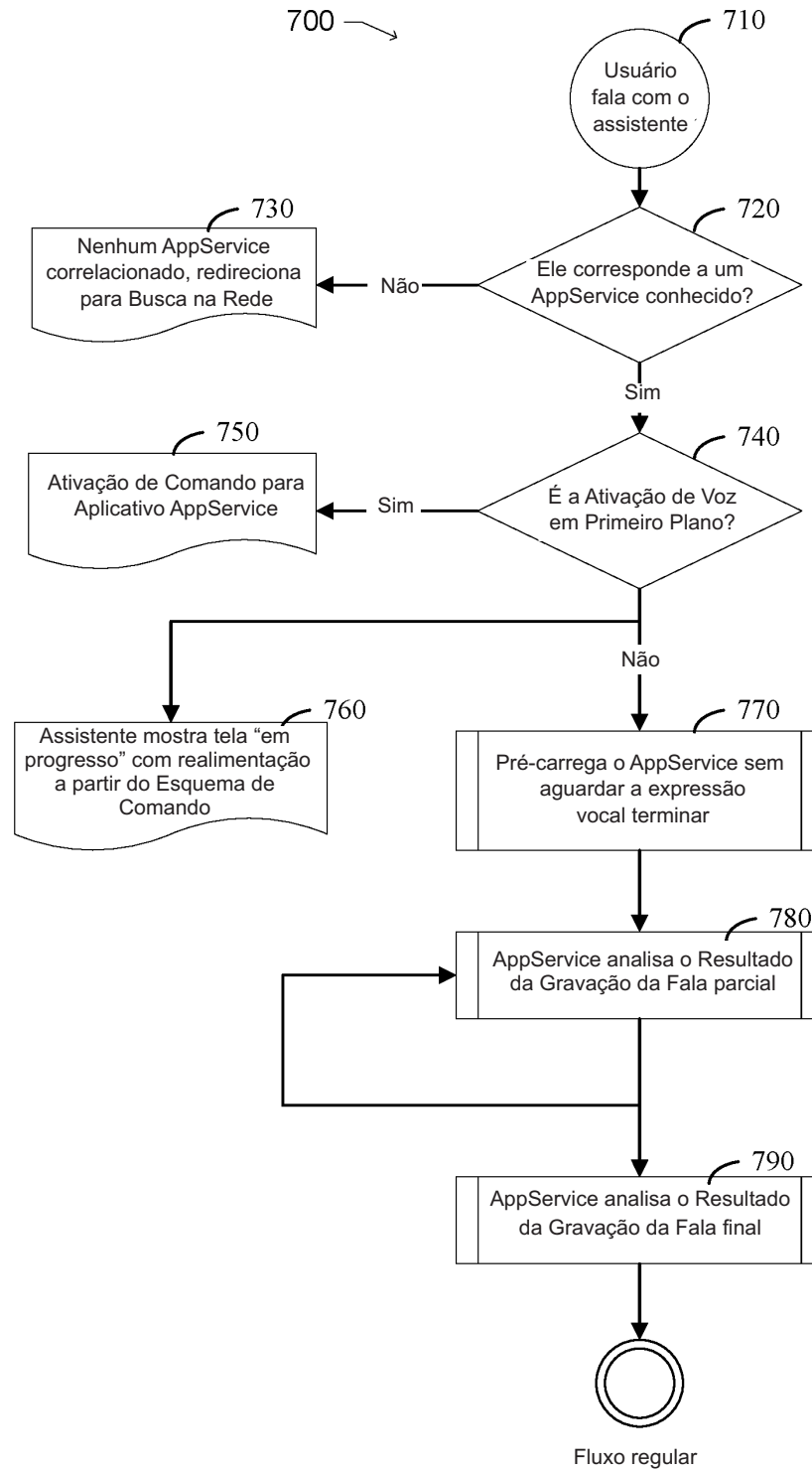




FIG. 8

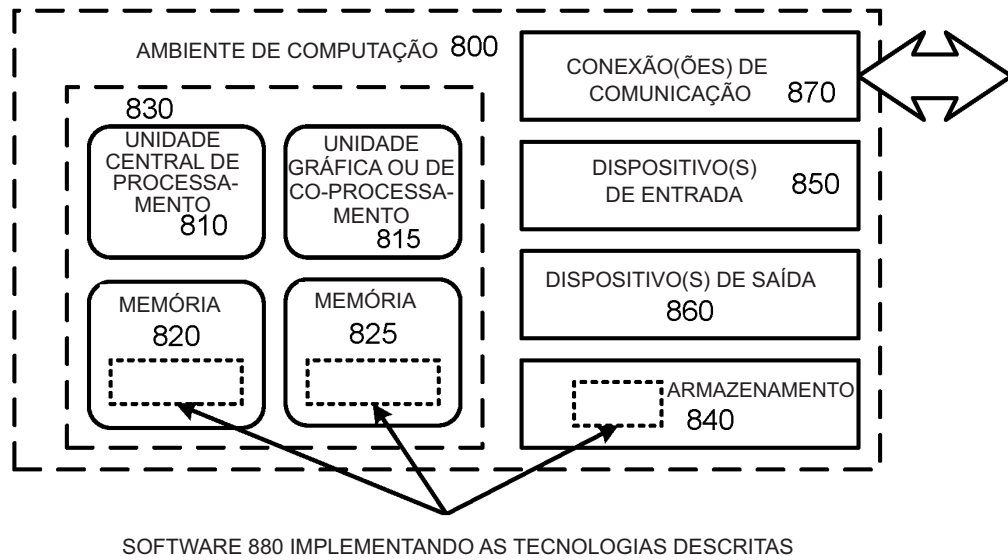


FIG. 9

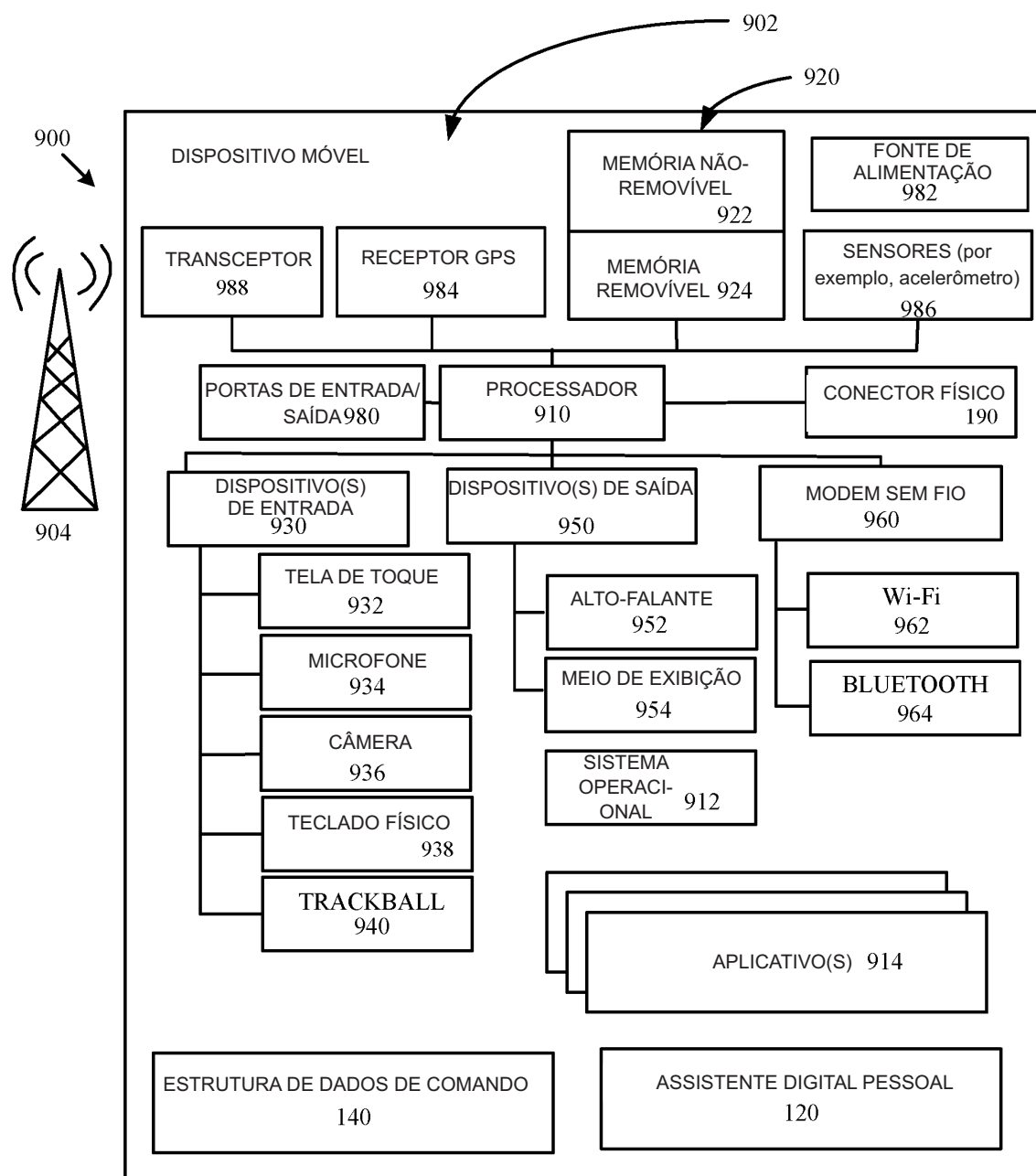


FIG. 10

