



República Federativa do Brasil
Ministério da Economia
Instituto Nacional da Propriedade Industrial

(11) BR 102014003975-9 B1



* B R 1 0 2 0 1 4 0 0 3 9 7 5 B 1 *

(22) Data do Depósito: 20/02/2014

(45) Data de Concessão: 03/03/2022

(54) Título: SISTEMA E MÉTODO PARA DETECÇÃO DE PALAVRA-CHAVE EM DISCURSO DE UM DISPOSITIVO

(51) Int.Cl.: G10L 15/00.

(52) CPC: G10L 15/00.

(30) Prioridade Unionista: 12/09/2013 CN 201310659840.4.

(73) Titular(es): MEDIATEK INC..

(72) Inventor(es): CHAO-LING HSU; YIOU-WEN CHENG; LIANG-CHE SUN; YUANYUAN WANG.

(57) Resumo: SISTEMA PARA DETECÇÃO DE PALAVRA CHAVE NO DISCURSO E MÉTODO ASSOCIADO. A invenção fornece um sistema para detecção de palavra chave no discurso e método associado. O sistema inclui um detector de palavra chave no discurso, um previsor de atividade e um tomador de decisão. O previsor de atividade obtém dados do sensor fornecidos por uma pluralidade de sensores, e processa os dados do sensor para fornecer um resultado de previsão de atividade indicando uma probabilidade de se um usuário está prestes a emitir uma palavra chave por voz. O tomador de decisão processa o resultado de previsão de atividade e um resultado de detecção de palavra chave preliminar da detecção de palavra chave no discurso para fornecer um resultado de detecção de palavra chave.

**"SISTEMA E MÉTODO PARA DETECÇÃO DE PALAVRA-CHAVE EM DISCURSO
DE UM DISPOSITIVO"**

CAMPO DA INVENÇÃO

[0001] A invenção refere-se a um sistema para detecção de palavra-chave em discurso e método associado, mais particularmente, a um sistema para melhorar a detecção de palavra-chave em discurso pela exploração de sensores para detectar a atividade do usuário, e método associado.

FUNDAMENTO DA INVENÇÃO

[0002] Dispositivos eletrônicos com controle de voz e/ou capacidade de interação tornam-se cada vez mais populares, porque eles podem fornecer interface de usuário à mão livre. O reconhecimento de voz para identificar palavras-chave, incluindo os comandos, na voz é essencial para implementar o controle de voz e/ou capacidade de interação.

SUMÁRIO DA INVENÇÃO

[0003] A invenção descreve um sistema para detecção de palavra-chave em discurso, incluindo um detector de palavra-chave em discurso, um previsor de atividade, um tomador de decisão, uma base de dados de atividade e uma base de dados de palavra-chave. A base de dados de atividade inclui um número de (uma ou mais) listas de atividade; cada lista de atividade é associada com um aplicativo alvo, e inclui um ou mais modelos de atividade. A base de dados de palavra-chave inclui uma ou mais listas de palavra-chave; cada lista de palavra-chave é associada com um aplicativo alvo, e inclui uma ou mais palavras chaves candidatas.

[0004] O previsor de atividade obtém dados do sensor fornecidos por um ou mais sensores, obtém a lista de atividade selecionada a partir da base de dados de atividade

com o aplicativo alvo da atividade selecionada correspondendo a um aplicativo atualmente em execução do dispositivo, e dessa forma processando os dados do sensor para fornecer um resultado de previsão de atividade indicando uma probabilidade de se um usuário está prestes a emitir a palavra-chave por voz. O previsor de atividade compara os dados do sensor com cada modelo da atividade da lista de atividade selecionada, e consequentemente, fornece um resultado correspondente de atividade para ser incluído no resultado de previsão de atividade. Alternativamente, o previsor de atividade obtém dados extraídos pelos recursos de extração dos dados do sensor, e então compara os dados extraídos com cada modelo da atividade da lista de atividade selecionada para fornecer consequentemente um resultado correspondente de atividade para ser incluído no resultado de previsão de atividade.

[0005] O detector de palavra-chave em discurso obtém uma lista de palavra-chave selecionada a partir da base de dados de palavra-chave com o aplicativo alvo da lista de palavra-chave selecionada correspondendo ao aplicativo atualmente em execução do dispositivo, e dessa forma fornecendo um resultado de detecção de palavra-chave preliminar. De preferência, o detector de palavra-chave em discurso compara o som que entra com cada palavra-chave candidata da lista de palavra-chave selecionada para dessa forma fornecer o resultado de detecção de palavra-chave preliminar.

[0006] O tomador de decisão é acoplado ao previsor de atividade e o detector de palavra-chave em discurso, e é capaz de processar o resultado de previsão de atividade e o resultado de detecção de palavra-chave preliminar para fornecer um segundo (exemplo, um aprimorado) resultado de detecção de palavra-chave.

Por exemplo, o tomador de decisão pode adotar um algoritmo de fusão de probabilidade com base em, por exemplo, teoria de Dempster-Shafer ou um algoritmo de aprendizagem de máquina com base em, por exemplo, modelo de mistura Gaussiano para processar o resultado de previsão de atividade e o resultado de detecção de palavra-chave preliminar, e assim fornecer o segundo resultado de detecção de palavra-chave. Para outro exemplo, o tomador de decisão pode calcular uma combinação linear (por exemplo, uma soma ponderada) do resultado de previsão de atividade e o resultado de detecção de palavra-chave preliminar como o segundo resultado de detecção de palavra-chave.

[0007] O sistema inclui ainda um detector de voz. O detector de voz é acoplado ao detector de palavra-chave em discurso, e é capaz de avaliar a informatividade (por exemplo, pela SNR, razão sinal para ruído) do som que entra, e dessa forma determinar quando ativar o detector de palavra-chave em discurso. Quando a informatividade do som que entra é baixa (por exemplo, abaixo de um limiar de informatividade), o detector de voz desativa o detector de palavra-chave em discurso. Quando a informatividade do som que entra é alta (por exemplo, maior do que o limiar de informatividade), o detector de voz ativa o detector de palavra-chave em discurso. Quando o detector de palavra-chave em discurso é desativado, o previsor de atividade e/ou o tomador de decisão não tem que trabalhar, assim o previsor de atividade e/ou o tomador de decisão também pode ser desativado. Quando o detector de palavra-chave em discurso é ativado, o previsor de atividade e/ou o tomador de decisão também pode ser ativado para cooperar com o detector de palavra-chave em discurso.

[0008] O detector de voz inclui um primeiro

estimador, um segundo estimador e um comparador acoplado ao primeiro estimador, o segundo estimador e o detector de palavra-chave em discurso. O primeiro estimador gera uma amostra atual de uma primeira sequência como uma soma ponderada de um volume atual do som que entra e uma amostra anterior da primeira sequência; isto é, computando a amostra atual da primeira sequência pela soma de um primeiro produto e um segundo produto, sendo que o primeiro produto é um resultado da multiplicação da amostra anterior da primeira sequência por um primeiro peso, e o segundo produto é um resultado da multiplicação da amostra atual do som que entra por um primeiro peso de complementaridade. O primeiro peso pode ser maior que 0 e menor que 1; o primeiro peso complementar pode ser igual a 1 menos o primeiro peso.

[0009] O segundo estimador gera uma amostra atual de uma segunda sequência como uma soma ponderada do volume atual do som que entra e uma amostra anterior da segunda sequência; isto é, computando a amostra atual da segunda sequência pela soma de um terceiro produto e um quarto produto, sendo que o terceiro produto é um resultado da multiplicação da amostra anterior da segunda sequência por um segundo peso, e o quarto produto é um resultado da multiplicação da amostra atual do som que entra por um segundo peso de complementaridade. O segundo peso pode ser maior que 0 e menor que 1; o segundo peso complementar pode ser igual a 1 menos o segundo peso.

[0010] A invenção ainda divulga um método para a detecção de palavra-chave em discurso, incluindo: obter dados do sensor fornecidos por um ou mais sensores, dessa forma computar um resultado de previsão de atividade indicando uma probabilidade de se um usuário está prestes a emitir a palavra-chave por voz, e computar um segundo resultado de detecção de palavra-chave

de acordo com os resultados de previsão de atividade e um resultado de detecção de palavra-chave preliminar da detecção de palavra-chave em discurso.

[0011] Vários objetos, características e vantagens da invenção serão facilmente aparentes mediante a leitura da seguinte descrição detalhada das modalidades da invenção considerada em conjunto com os desenhos anexos. No entanto, os desenhos aqui empregados são para efeitos de descrição e não devem ser considerados como limitantes.

BREVE DESCRIÇÃO DOS DESENHOS

[0012] Os objetos acima e as vantagens da invenção irão tornar-se mais prontamente aparentes para aqueles normalmente versados na técnica após a revisão da descrição detalhada a seguir e desenhos de acompanhamento, em que:

[0013] FIG. 1 ilustra um sistema para detecção de palavra-chave em discurso de acordo com uma modalidade da invenção;

[0014] FIG. 2 ilustra curvas de dados do sensor dos resultados detectados e resultados detectados extraídos;

[0015] FIG. 3 ilustra um fluxograma para detecção de palavra-chave em discurso assistida por sensor de acordo com uma modalidade da invenção;

[0016] FIG. 4 a FIG. 7 ilustram exemplos de operação da detecção de palavra-chave em discurso assistida por sensor de acordo com uma modalidade da invenção;

[0017] FIG. 8 ilustra um fluxograma para modificar a base de dados de atividade mostrada na FIG. 1;

[0018] FIG. 9 ilustra um sistema para detecção de palavra-chave em discurso de acordo com uma modalidade da

invenção;

[0019] FIG. 10 ilustra uma modalidade do detector de voz mostrada na FIG. 9; e

[0020] FIG. 11 ilustra um fluxograma para detecção de palavra-chave em discurso assistida por sensor de baixa energia.

DESCRIÇÃO DETALHADA DAS MODALIDADES PREFERIDAS

[0021] Para aumentar a conscientização e interatividade com o usuário e arredores, o dispositivo eletrônico moderno está equipado com um grande número de sensores. Por exemplo, os sensores de um dispositivo de comunicação, por exemplo, um telefone móvel inteligente, pode incluir um acelerômetro, um giroscópio, um magnetômetro, um barômetro, um sensor de proximidade, sensor de luz, uma tela sensível ao toque, um receptor de um sistema de posicionamento, um receptor sem fio e/ou uma câmera etc.

[0022] Dados do sensor fornecidos pelo sensor (es) de um dispositivo podem ser aproveitados para obter informações de atividade sobre a interação usuário-dispositivo, status do usuário e/ou do ambiente em torno do dispositivo. A informação de atividade sobre a interação usuário-dispositivo pode incluir: (a) se o dispositivo é elevado, abaixado, levantado, colocado no chão, virado, deixado cair, agitado, constantemente mantido, inclinado, mantido plano, aproximado a alguma coisa (por exemplo, usuário), afastado de alguma coisa, e/ou colocado em ambiente escuro (por exemplo, em uma bolsa ou mochila) ou ambiente iluminado etc.; (b) um evento de entrada que representa que o dispositivo precisa interagir com o usuário, por exemplo, se o dispositivo recebe uma chamada, mensagem e/ou e-mail, e/ou se o dispositivo for alarmar em um momento predefinido, tal como

uma chamada de manhã, uma chamada para despertar, um alarme, um lembrete, uma pop-up da tela para entrada de item a fazer, reunião listada no calendário, agenda e/ou cronograma. A informação de atividade sobre o status do usuário pode incluir se o usuário está andando, correndo e/ou dirigindo, etc. A informação de atividade sobre o ambiente pode incluir: temperatura ambiente, ruído, luminosidade, localização, posição e/ou hora atual.

[0023] Nesta modalidade, a palavra-chave por voz do usuário (comando) muitas vezes ocorre após (e/ou junto com) a atividade de padrão reconhecível. Tomando um telefone celular como um exemplo: o telefone toca para uma chamada recebida, a resposta natural do usuário é levantar primeiro o telefone para ver informações relacionadas, por exemplo, a identificação de quem é a chamada, e, em seguida, decide como responder, por exemplo, para responder ou ignorar/rejeitar a chamada. Assim, a atividade de captação é uma pista para sinalizar para o telefone esperar palavra-chave atender por voz (comando). Alternativamente, quando o usuário deseja tirar uma foto pela função de câmera do telefone, a ação natural do usuário é primeiro manter o telefone fixo, e, em seguida, instruir o telefone a tirar a foto. Assim, a atividade de manter estável fornece informações sobre quando esperar uma palavra-chave disparar por voz.

[0024] De preferência, o dado do sensor é utilizado para indicar se a atividade do padrão conhecido ocorre, e aumenta de acordo com a detecção da palavra-chave em discurso, pelo fornecimento de informações adicionais, por exemplo, prevendo quando o usuário vai dizer a palavra-chave por voz. Por exemplo, um palavra-chave pode ser um comando, uma instrução, um termo para a consulta em mecanismos de pesquisa e/ou mecanismos de

inteligência artificial, e/ou uma voz informativa, por exemplo, "Yee-Ha!", embora esta possa não ser um vocabulário oficial.

[0025] Por exemplo, quando o telefone toca para uma atividade de chamada recebida e a atividade de levantar o telefone é detectada, o telefone pode, por conseguinte prever que o usuário vai dizer a palavra-chave por voz como "atender" ou "rejeitar", e, portanto, ajustar a sensibilidade da detecção da palavra-chave em discurso para facilitar o reconhecimento da palavra-chave para atender falada depois, por exemplo, "atender" ou "rejeitar". Por exemplo, quando o telefone estiver ligado a função de câmera e a atividade de manter estável for detectada, o telefone pode esperar uma palavra-chave por voz para disparar, por exemplo, "xis", para tirar a foto, e em seguida, ajustar a sensibilidade da detecção da palavra-chave em discurso para facilitar o reconhecimento da palavra-chave disparar por voz.

[0026] Assim, detecção de palavra-chave em discurso pode ser aprimorada de acordo com a previsão da invenção, sendo que a previsão de atividade é projetada para aproveitar dados do sensor e consequentemente detectar a ocorrência de atividades que indicam, o que acontece antes (ou quando) o usuário está prestes a dizer a palavra-chave por voz. Além disso, detecção de palavra-chave em discurso e previsão de atividade pode ser realizada no contexto de cenário de aplicação. Nesta modalidade, quando um telefone está executando um aplicativo de comunicação para tocar para uma chamada recebida, a previsão de atividade é arranjada para detectar a ocorrência de um primeiro conjunto de atividades indicativas relacionadas (por exemplo, elevação do telefone), e detecção de palavra-chave em discurso é arranjada para reconhecer um primeiro conjunto de palavras chaves por voz.

relacionadas, como palavras chaves para atender, por exemplo, "atender" ou "rejeitar". Quando o telefone está executando um aplicativo de câmera, a previsão de atividade é arranjada para detectar a ocorrência de um segundo conjunto de atividades indicativas relacionadas (por exemplo, manter estável), e detecção de palavra-chave em discurso é arranjada para reconhecer um segundo conjunto de palavras chaves por voz relacionadas, por exemplo, a palavra-chave de disparo por voz como "xis".

[0027] Há dois tipos de erros para degradar o desempenho da detecção de palavra-chave em discurso, incluindo erro de falta e erro de alarme falso. Erro de falta acontece quando o usuário diz a palavra-chave por voz, mas a detecção de palavra-chave em discurso falha em reconhecer a palavra-chave por voz dita. Erro de alarme falso acontece quando o usuário não diz qualquer palavra-chave por voz, mas a detecção de palavra-chave em discurso erroneamente determina que uma palavra-chave por voz foi dita.

[0028] Detecção de palavra-chave em discurso tem uma sensibilidade ajustável (ou tendência de reconhecimento) para uma troca entre erro de falta e erro de alarme falso. Aumentar a sensibilidade faz a detecção de palavra-chave em discurso tender a interpretar o som que entra da palavra-chave por voz, mesmo quando o som que entra é menos provável de conter a palavra-chave por voz. Consequentemente, aumentar a sensibilidade reduz a probabilidade de erro de falta enquanto a probabilidade de erro de alarme falso é aumentada. Por outro lado, diminuir a sensibilidade diminui a tendência para a detecção de palavra-chave em discurso interpretar o som que entra da palavra-chave por voz, mesmo quando o som que entra é muito suscetível de conter a palavra-chave por voz. Por isso, a

diminuição da sensibilidade aumenta a probabilidade do erro de falta, mas reduz a probabilidade de erro de alarme falso.

[0029] Nesta modalidade, a sensibilidade da detecção da palavra-chave em discurso é adaptativamente e dinamicamente ajustada de acordo com a previsão de atividade, assim tanto o erro de falta quanto o erro de alarme falso pode ser suprimido sem comprometer. Quando uma atividade indicativa é detectada, a sensibilidade de reconhecer a palavra-chave por voz relacionada pode ser aumentada, assim o som que entra é mais suscetível a ser reconhecido como palavra-chave por voz relacionada, mesmo que a palavra-chave falada seja fraca, pouco clara e/ou ruidosa, daí, o erro de falta é suprimido. Por outro lado, durante a ausência de atividades indicativas, a sensibilidade da detecção de palavra-chave em discurso pode ser reduzida, assim o erro de alarme falso pode ser suprimido porque o som que entra é menos suscetível de ser incorretamente reconhecido como palavra-chave por voz.

[0030] Favor consultar a FIG. 1. FIG. 1 ilustra um dispositivo 10 integrado com um sistema 12 para melhorar um detector de palavra-chave em discurso 14 de acordo com uma modalidade da invenção. O dispositivo 10 pode ser um eletrônico consumidor, como um dispositivo de comunicação (por exemplo, um telefone móvel, um telefone celular, um smartphone), um computador de mão ou portátil, um dispositivo de cuidado pessoal acusticamente interativo e/ou controlável, brinquedo, dispositivo para vestir (por exemplo, relógio), televisão, reproduutor de mídia, gravador de mídia (por exemplo, câmera digital ou filmadora), dispositivo navegador ou de posicionamento etc.. Em uma modalidade, o detector da palavra-chave em discurso 14 é arranjado para manter

automaticamente o monitoramento se o som que entra contém a palavra-chave por voz mesmo sem sugestão do usuário.

[0031] Para aprimorar o desempenho do detector de palavra-chave em discurso 14, o sistema 12 inclui ainda uma base de dados de palavra-chave 16, um previsor de atividade 18, uma base de dados de atividade 20 e um tomador de decisão 22. A base de dados de palavra-chave 16 é acoplada ao detector de palavra-chave em discurso 14, e inclui um número N2 (igual a ou maior que 1) das listas de palavra-chave KL[1] a KL[N2]. De preferência, cada lista de palavra-chave KL[j] (para $j = 1$ a N2) é associada com um aplicativo alvo app[j], e inclui uma quantidade $P\{j\}$ (igual a ou maior que 1) de palavras chaves candidatas $kw[j,1]$ a $kw[j,P\{j\}]$. Diferentes listas de palavra-chave podem se associar com diferentes aplicativos alvos, e podem ter diferentes quantidades de palavras chaves candidatas. Isto é, para índices $j1$ não igual a $j2$, o aplicativo alvo app[j1] da lista de palavra-chave KL[j1] pode diferir do aplicativo alvo app[j2] da lista de palavra-chave KL[j2]; a quantidade $P\{j1\}$ da lista de palavra-chave KL[j1] pode diferir de ou ser igual a uma quantidade $P\{j2\}$ da lista de palavra-chave KL[j2].

[0032] A base de dados de atividade 20 é acoplada ao previsor de atividade 18, e inclui um número N1 (igual a ou maior que 1) de listas de atividade AL[1] a AL[N1]; cada lista de atividade AL[i] (para $i = 1$ a N1) é associada com um aplicativo alvo app[i], e inclui uma quantidade $Q\{i\}$ (igual a ou maior que 1) de modelos de atividade $at[i,1]$ a $at[i,Q\{i\}]$. Diferentes listas de atividade podem se associar com diferentes aplicativos alvos, e podem ter diferentes quantidades de modelos de atividade.

[0033] O detector de palavra-chave em discurso 14

recebe um sinal de som Snd. Por exemplo, o dispositivo 10 pode incluir microfone(s) e/ou conjunto(s) de microfone (não mostrado) para coletar o som, e dessa forma fornecer o sinal digital Snd pelo processamento (por exemplo, conversão de analógico para digital) do som coletado. Alternativamente, a partir de outro aparelho remoto (por exemplo, um microfone sem fio, não mostrado), o dispositivo 10 pode receber um sinal remotamente fornecido (não mostrado) que contém som codificado ou não codificado, e dessa forma fornecer o sinal de som Snd pelo processamento do sinal remotamente fornecido.

[0034] De acordo com um aplicativo atualmente em execução do dispositivo 10, o detector de palavra-chave em discurso 14 também pode obter uma lista de palavra-chave selecionada KL[jx] das listas de palavra-chave KL[1] a KL[N2] da base de dados de palavra-chave 16, em que o aplicativo alvo app[jx] da lista de palavra-chave selecionada KL[jx] corresponde ao aplicativo atualmente em execução do dispositivo 10. Por exemplo, do dispositivo 10 e/ou um sistema operacional (OS) do dispositivo 10, o detector de palavra-chave em discurso 14 e/ou a base de dados de palavra-chave 16 pode acessar um status que indica o aplicativo atualmente em execução do dispositivo 10, e pode então achar (selecionar) a lista de palavra-chave KL[jx] cujo aplicativo alvo app[jx] é igual ao aplicativo atualmente em execução do dispositivo 10. Aplicativos executados pelo dispositivo 10 podem se referir a programas utilitários, serviços, procedimentos e/ou sub-rotinas executados sob o controle do OS. Um aplicativo atualmente em execução pode referir-se a um aplicativo ativo no momento, um aplicativo de primeiro plano, uma aplicação de fundo, e/ou um aplicativo em foco.

[0035] Pela seleção da lista de palavra-chave

correspondente ao aplicativo atualmente em execução, a detecção de palavra-chave em discurso pode ser realizada com referência a um contexto adaptativamente. Por exemplo, a respeito de um aplicativo de comunicação que é responsável pela manipulação da chamada recebida, uma lista de palavra-chave correspondente pode incluir candidatos a palavras-chave para atender como "atender" e "rejeitar". Para um aplicativo de câmera, sua lista de palavra-chave correspondente pode incluir candidatos de palavra-chave para disparar como "xis".

[0036] Em resposta ao sinal Snd, o detector de palavra-chave em discurso 14 pode fornecer um resultado de detecção de palavra-chave preliminar Skw de acordo com a lista de palavra-chave selecionada KL[jx]. Por exemplo, o detector de palavra-chave em discurso 14 pode comparar o som que entra no sinal Snd com cada uma das palavras chaves candidatas kw[jx,1] a kw[jx,P{jx}] da lista de palavra-chave selecionada KL[jx] para fornecer o resultado de detecção de palavra-chave preliminar Skw.

[0037] O previsor de atividade 18 recebe um sinal de dados do sensor Ssd fornecido pelo sensor(es) do dispositivo 10. Por exemplo, o dispositivo 10 pode incluir sensor(es) para detectar o movimento, aceleração, localização, posição, direção e/ou atitude angular (por exemplo, sendo virado ou inclinado), volume circundante, brilho e/ou campo(s) de força exercido sobre o dispositivo 10 (por exemplo, campo magnético, eletromagnético e/ou de gravidade) como o sinal Ssd. Alternativamente, de outro aparelho remoto (por exemplo, sensor(es) remoto, não mostrado), o dispositivo 10 pode receber um sinal remotamente fornecido (não mostrado) que contém, incorpora, e/ou codificado com dados do sensor, e dessa forma fornecer o sinal Ssd pelo processamento

do sinal remotamente fornecido.

[0038] De acordo com os aplicativo atualmente em execução do dispositivo 10, o previsor de atividade 18 obtém a lista de atividade selecionada AL[ix] das listas de atividade AL[1] a AL[N1] da base de dados de atividade 20, em que o aplicativo alvo app[ix] da lista de atividade selecionada AL[ix] representa o aplicativo atualmente em execução do dispositivo 10. Por exemplo, do dispositivo 10 e/ou OS do dispositivo 10, o previsor de atividade 18 e/ou a base de dados de atividade 20 obtém um status indicando o aplicativo atualmente em execução do dispositivo 10, e então seleciona a lista de atividade AL[ix] associada com um aplicativo alvo app[ix] indicando o aplicativo atualmente em execução do dispositivo 10. Pela seleção da lista de atividade associada com um aplicativo atualmente em execução, a previsão de atividade pode ser realizada de uma maneira adaptativa ao contexto. Por exemplo, a respeito de uma aplicação de comunicação responsável pela manipulação de chamada recebida, uma lista de atividade correspondente pode incluir um modelo de atividade da captação do telefone, para um aplicativo de câmera, a sua lista de atividades correspondente pode incluir um modelo de atividade para se manter constante.

[0039] De acordo com os lista de atividade selecionada AL[ix], o previsor de atividade 18 processa o sinal Ssd para fornecer um resultado de previsão de atividade Sap indicando uma probabilidade de se um usuário estiver prestes a emitir uma palavra-chave por voz. Por exemplo, o previsor de atividade 18 compara o sinal Ssd com cada um dos modelos de atividade at[ix,1] a at[ix,Q{ix}] registrados na lista de atividade selecionada AL[ix], e dessa forma fornece um resultado correspondente de atividade como o resultado de previsão de

atividade Sap.

[0040] Em uma modalidade, cada modelo de atividade $at[i,q]$ pode incluir resultado(s) padrão, típico, representativo e/ou mais frequentemente detectado de uma atividade indicativa (movimento ou estado) que acontece antes ou quando o usuário está prestes a dizer a palavra-chave por voz. Cada resultado detectado se associa com um sensor e é registrado como uma referência no modelo de atividade $at[i,q]$. Quando o previsor de atividade 18 gera o resultado Sap pela comparação dos dados do sensor Ssd com cada modelo de atividade $at[ix,q]$ da lista de atividade selecionada $AL[ix]$, para cada resultado detectado de um dado tipo de sensor incluído no sinal de dados do sensor Ssd, por exemplo, o previsor de atividade 18 checa se o modelo de atividade $at[ix,q]$ inclui uma referência associada com um mesmo tipo de sensor; se verdadeiro, o previsor de atividade 18 compara o resultado detectado e a referência respectivamente incluída no sinal Ssd e o modelo de atividade $at[ix,q]$ no mesmo tipo de sensor, e então reflete o resultado da comparação no sinal Sap.

[0041] Em uma modalidade, cada modelo de atividade $at[i,q]$ inclui referências extraídas, cada referência extraída é associada com um sensor, e representa as feições extraídas de um resultado detectado de uma atividade indicativa. Quando o previsor de atividade 18 gera o resultado Sap pela comparação do sinal de dados do sensor Ssd com cada modelo de atividade $at[ix,q]$ da lista de atividade selecionada $AL[ix]$, o previsor de atividade 18 pode primeiro extrair as feições de cada resultado detectado incluído no sinal de dados do sensor Ssd para dessa forma gerar um resultado detectado extraído (não mostrado); para cada resultado detectado extraído de um dado tipo de sensor

incluído no sinal Ssd, o previsor de atividade 18 pode então achar quando o modelo de atividade at[ix,q] contém uma referência extraída para o mesmo tipo do sensor; se verdadeiro, o previsor de atividade 18 compara o resultado detectado extraído e a referência extraída respectivamente incluída no sinal Ssd e o modelo de atividade at[ix,q] no mesmo tipo de sensor, e então reflete o resultado da comparação no sinal Sap.

[0042] A extração das feições de um resultado detectado pode ser conseguida pela filtragem (por exemplo, filtragem passa baixa) do resultado detectado, cálculo das estatísticas do resultado detectado, e/ou transformação do resultado detectado em domínio de espectro. Favor consultar a FIG. 2 que ilustra exemplos dos resultados detectados e resultados detectados extraídos; os resultados detectados são detectados por um acelerômetro. Como mostrado na FIG. 2, as curvas de dados do sensor s1x, s1y e s1z representam, respectivamente, eixo x, eixo y e eixo z da aceleração detectada durante uma atividade de inversão, e formam em conjunto um resultado detectado do acelerômetro. A extração das feições (por exemplo, filtragem passa baixa) das curvas s1x, s1y e s1z pode respectivamente obter três curvas de dados do sensor extraídas e1x, e1y e e1z para coletivamente formar um resultado detectado extraído. Similarmente, as curvas de dados do sensor s2x, s2y e s2z representam, respectivamente, o eixo x, eixo y e eixo z da aceleração detectada durante uma atividade de manter estável. Pela extração dos recursos das curvas s2x, s2y e s2z, três curvas de dados do sensor extraídas e2x, e2y e e2z podem ser obtidas respectivamente. Características de cada resultado detectado e resultado detectado extraído podem ser usadas para distinguir diferentes atividades. Por exemplo, uma transição a partir das

amplitudes de +9,8 a -9,8 pode ser observada na curva slz ou a curva elz; portanto, características como a duração da transição (por exemplo, intervalo Td ou Te), tempo de cruzamento zero e/ou diferença de amplitude da transição podem ser utilizadas para distinguir se o dispositivo é deixado cair ou virado pelo usuário.

[0043] Favor consultar a FIG. 1 novamente. O tomador de decisão 22 é acoplado ao previsor de atividade 18 e o detector de palavra-chave em discurso 14, e é capaz de processar o resultado de previsão de atividade Sap e o resultado de detecção de palavra-chave preliminar Skw para fornecer um segundo resultado de detecção de palavra-chave Sdm, assim o dispositivo 10 pode responder de acordo com o resultado Sdm. Porque o resultado Sdm é atingido pelo processamento do resultado de previsão de atividade Sap e o resultado de detecção de palavra-chave do discurso original Skw, o resultado Sdm é mais confiável do que o resultado Skw, e assim a detecção de palavra-chave em discurso do detector de palavra-chave em discurso 14 é aprimorada.

[0044] Junto com a FIG. 1, favor consultar a FIG. 3 que ilustra um fluxograma 100 de acordo com uma modalidade da invenção. O sistema 10 na FIG. 1 pode adotar o fluxograma 100 para aprimorar o desempenho do detector de palavra-chave em discurso 14. O fluxograma 100 inclui as etapas a seguir.

[0045] Etapa 102: identificar o aplicativo atualmente em execução do dispositivo 10. Como mencionado anteriormente, o sistema 12 pode acessar um status do dispositivo 10 para identificar o aplicativo atualmente em execução. Por exemplo, o status pode ser fornecido pelo OS do dispositivo 10, e/ou por um registro de uma CPU (unidade de processamento central, não mostrada) que controla o dispositivo 10.

[0046] Etapa 104: selecionar uma lista de atividade correspondente AL[ix] e uma lista de palavra-chave correspondente KL[jx] respectivamente da base de dados de atividade 20 e a base de dados de palavra-chave 16.

[0047] Etapa 106: pelo o detector de palavra-chave em discurso 14, realizar uma detecção de palavra-chave em discurso preliminar com base no sinal de som Snd e a lista de palavra-chave selecionada KL[ix], de modo a fornecer o resultado de detecção de palavra-chave preliminar Skw. Por exemplo, o detector de palavra-chave em discurso 14 pode comparar o som no sinal Snd (FIG. 1) com as palavras chaves candidatas kw[jx,1] a kw[jx,P{jx}] para respectivamente obter resultados de comparação scr[1] a scr[P{jx}], cada resultado scr[p] (para p=1 a P{jx}) quantitativamente indica a similaridade entre o sinal Snd e a palavra-chave candidata kw[jx,p], e/ou indica a probabilidade ou a possibilidade de que o sinal Snd corresponda à palavra-chave candidata kw[jx,p]. O detector de palavra-chave em discurso 14 também pode obter um resultado de comparação adicional scr[P{jx}+1] para quantitativamente indicar a probabilidade ou possibilidade de que o sinal Snd falhe em corresponder a qualquer das palavras chaves candidatas kw[jx,1] a kw[jx,P{jx}], isto é, para indicar a probabilidade para o sinal Snd ser “palavra não-chave”. Então o detector de palavra-chave em discurso 14 pode encontrar uma máxima dos resultados scr[1] a scr[P{jx}+1] e assim fornecer o resultado Skw. Para conveniência da discussão, o máximo dos resultados scr[1] a scr[P{jx}+1] é denotado como o resultado scr[p_max] com índice p_max sendo um dos números 1 a (P{jx}+1), e a palavra-chave candidata kw[jx,p_max] correspondendo ao resultado scr[p_max] é referida como uma palavra-chave mais provável; se o índice

`p_max` é igual a $(P\{jx\}+1)$, a palavra-chave mais provável `kw[jx,p_max]` refere-se à situação de “palavra não-chave”.

[0048] Para obter os resultados `scr[1]` a `scr[P{jx}+1]`, o detector de palavra-chave em discurso 14 (FIG. 1) pode ser um mapeador de palavra-chave que adota algoritmo(s) de processamento de discurso com base em, por exemplo, modelo de Markov escondido (HMM) etc.. Cada um dos resultados `scr[1]` a `scr[P{jx}+1]` pode ser uma classificação linear ou uma probabilidade de logaritmo.

[0049] Etapa 108: de acordo com a lista de atividade selecionada `AL[ix]` e dados do sensor no sinal de dados do sensor `Ssd`, compute o resultado de previsão de atividade `Sap` pelo previsor de atividade 18 (FIG. 1) para indicar uma probabilidade ou possibilidade para quando um usuário está prestes a emitir uma palavra-chave por voz. Por exemplo, o previsor de atividade 18 compara o sinal `Ssd` com cada modelo de atividade `at[ix,q]` da lista de atividade selecionada `AL[ix]` (para $q=1$ a $Q\{ix\}$), e dessa forma fornecer um resultado que corresponde a atividade `acr[q]` para quantitativamente indicar similaridade entre o modelo de atividade `at[ix,q]` e o sinal `Ssd`, e/ou probabilidade ou possibilidade de que o sinal `Ssd` corresponda ao modelo de atividade `at[ix,q]`. Os resultados `acr[1]` a `acr[Q\{ix\}]` são então coletados para formar o resultado `Sap`, por exemplo, o resultado `Sap` inclui os resultados `acr[1]` a `acr[Q\{ix\}]`. Porque o sinal de dados do sensor `Ssd` pode incluir os resultados detectados de diferentes sensores (ou pode fornecer resultados detectados extraídos de diferentes sensores), e cada modelo de atividade `at[ix,q]` pode incluir referências (ou referências extraídas) de diferentes sensores, o previsor de atividade 18 pode comparar cada resultado detectado (ou cada resultado detectado extraído)

de um sensor com uma referência (ou uma referência extraída) do mesmo sensor para obter um resultado de comparação individual, e gerar o resultado correspondente acr[q] de acordo com os resultados de comparação de diferentes sensores.

[0050] Para obter os resultados acr[1] a acr[Q{ix}], o previsor de atividade 18 pode adotar o algoritmo baseado em regras, ou algoritmo(s) mais sofisticado com base no modelo de mistura Gaussiano, modelo de Markov escondido, máquina de vetor de suporte e/ou rede neural etc.. Alternativamente, o previsor de atividade 18 pode adotar similarmente algoritmo(s) de medição com base no distorção de tempo dinâmica etc.. Observe que as etapas 106 e 108 podem ser executadas simultaneamente ou em ordem sequencial.

[0051] Etapa 110: pelo tomador de decisão 22 (FIG. 1), faça uma decisão final para fornecer o resultado de detecção de palavra-chave Sdm de acordo com o resultado de detecção de palavra-chave preliminar Skw e o resultado de previsão de atividade Sap. O resultado Sdm pode ser uma probabilidade, uma pontuação e/ou uma logaritmo medindo quantitativamente o quanto provável a palavra-chave mais provável kw[jx,p_max] da etapa 106 ser realmente dita no contexto do aplicativo atualmente em execução e atividade do usuário.

[0052] O dispositivo 10 pode periodicamente repetir o fluxograma 100 para realizar a detecção de palavra-chave em discurso assistido por sensor. Alternativamente, o dispositivo 10 pode executar o fluxograma 100 quando necessário, por exemplo, quando o usuário instrui.

[0053] Em uma modalidade, o resultado Sdm é obtido pela verificação se o resultado Skw satisfaz uma primeira condição e o resultado Sap satisfaz uma segunda condição. Por

exemplo, a primeira condição pode ser satisfeita se o resultado Skw for maior do que o primeiro limiar, e a segunda condição pode ser satisfeita se cada um dos resultados acr[1] a acr[Q{ix}] no resultado Sap for maior do que um segundo limiar. Alternativamente, a segunda condição é: se uma soma (ou uma combinação linear) dos resultados acr[1] a acr[Q{ix}] for maior do que um segundo limiar. Alternativamente, a segunda condição é: se uma propriedade estatística (por exemplo, máxima, mínima, média etc.) dos resultados acr[1] a acr[Q{ix}] for maior do que um segundo limiar. De preferência, quando tanto a primeira quanto a segunda condição são satisfeitas, o tomador de decisão 22 (FIG. 1) pode determinar que a palavra-chave mais provável kw[jx,p_max] é dita e reflete a afirmativa com o resultado Sdm, assim o dispositivo 10 pode, portanto, responder, por exemplo, realizar operações associadas com uma palavra-chave kw[jx,p_max]. Por outro lado, se uma das primeiras e segunda condição não é satisfeita (ou a determinada palavra-chave kw[jx,p_max] representa “palavra não-chave”), então o tomador de decisão 22 pode refletir a negação com o resultado Sdm, assim a operação do dispositivo 10 não será afetada.

[0054] Em uma modalidade, o resultado Sdm é obtido pela computação de uma combinação linear dos resultados acr[1] a acr[Q{ix}] e Skw, e comparação se a combinação linear for maior do que um limiar predefinido; se verdadeiro, o tomador de decisão 22 determina que a palavra-chave mais provável kw[jx,p_max] é ouvida, caso contrário o tomador de decisão 22 determina que a palavra-chave kw[jx,p_max] não é reconhecida.

[0055] Em outras modalidades, o tomador de decisão 22 pode adotar um algoritmo de fusão de probabilidade com base em, por exemplo, teoria de Dempster-Shafer, ou um algoritmo de

aprendizagem de máquina com base em, por exemplo, modelo de mistura Gaussiano, para processar os resultados Skw e $acr[1]$ a $acr[Q\{ix\}]$ e dessa forma atingir um resultado mais confiável Sdm . Os algoritmos sofisticados acima mencionados podem aplicar número arbitrário de probabilidades como entradas e consequentemente fornecer uma probabilidade conclusiva como uma saída, por isso oferece uma solução flexível para integrar a informação respectivamente fornecida pelos resultados Skw e $acr[1]$ a $acr[Q\{ix\}]$, já que o número $Q\{ix\}$ pode ser diferente sob contextos de aplicação diferentes. Alternativamente, diferentes algoritmos podem ser combinados para gerar o resultado Sdm . Por exemplo, o tomador de decisão 22 adota um primeiro algoritmo para processar os resultados $acr[1]$ a $acr[Q\{ix\}]$ para dessa forma obter um primeiro resultado, e adota um segundo algoritmo para processar o primeiro resultado e o resultado Skw para assim obter o resultado Sdm .

[0056] Junto com a FIG. 1, favor consultar a FIG. 4 até a FIG. 7 que ilustram exemplos de operação do dispositivo 10 (FIG. 1). Assuma que o dispositivo 10 é um telefone móvel inteligente para a FIG. 4 a FIG. 7. A FIG. 4 demonstra um cenário de tirar foto. Quando o usuário quer tirar uma foto com o dispositivo 10, uma sequência natural, conveniente e ação intuitiva para o usuário é: ativar o aplicativo de câmera do dispositivo 10, manter o dispositivo 10 estável, e dizer a palavra-chave disparar por voz, por exemplo, "xis". Assim, para implementar um disparo de foto por controle de voz assistido por sensor, uma lista de atividade (FIG. 1) correspondente ao aplicativo de câmera pode incluir um modelo de atividade que representa a atividade indicativa de manter estável, e a lista de palavra-chave correspondente ao aplicativo de câmera pode

incluir a palavra(s) chave de disparo por voz candidata .

[0057] O disparo de foto por controle de voz assistido por sensor funciona como segue. Quando o usuário ativa o aplicativo de câmera do dispositivo 10 para preparar para disparar a foto, o previsor de atividade 18 (FIG. 1) do dispositivo 10 aproveita dados do sensor para detectar quando a atividade indicativa de manter estável ocorre. Quando o usuário está mantendo o dispositivo 10 estável, o previsor de atividade 18 refletirá a ocorrência da atividade indicativa, assim o dispositivo 10 pode prever que o usuário vai falar a palavra-chave de disparo por voz para disparar a captação da foto, e o dispositivo 10 permite que a palavra-chave de disparo por voz seja reconhecida mais facilmente. Por exemplo, o tomador de decisão 22 (FIG. 1) aumenta a sensibilidade da identificação da palavra-chave de disparo por voz, daqui a palavra-chave de disparo por voz pode ser reconhecida mesmo quando é dita fracamente ou em ambientes ruidosos. Uma vez que o tomador de decisão 22 reflete que a palavra-chave de disparo por voz é reconhecida, o dispositivo 10 pode responder tirando a foto. De modo contrário, quando a atividade indicativa de manter estável não é detectada, é improvável que o usuário diga a palavra-chave de disparo por voz, assim o dispositivo 10 pode evitar o reconhecimento errôneo da palavra-chave de disparo por voz. Por exemplo, o tomador de decisão 22 diminui a sensibilidade da identificação da palavra-chave de disparo por voz.

[0058] FIG. 5 demonstra um cenário de atender uma chamada recebida. Quando o dispositivo 10 recebe uma chamada recebida e toda para chamar a atenção do usuário, uma sequência de ação natural e amigável para o usuário é: elevar o dispositivo 10 para ver a informação sobre a chamada recebida (por exemplo,

quem está ligando), determinar como responder (por exemplo, atender a chamada, rejeitá-la ou ignorá-la), e consequentemente dizer uma palavra-chave para atender por voz, por exemplo, "atender," "rejeitar," "ignorar," ou "mudo." Nesta modalidade, para implementar uma resposta à chamada por controle de voz assistido por sensor, uma lista de atividade correspondente a um aplicativo de comunicação responsável por gerenciar a chamada recebida pode incluir um modelo de atividade que grava a atividade indicativa de levantar o telefone, e a lista de palavra-chave correspondente ao aplicativo de comunicação pode incluir as palavras chaves candidatas para atender por voz por exemplo, "atender," "rejeitar," "ignorar," ou "mudo." Por exemplo, quando o usuário está em uma reunião e mesmo fazendo um discurso, ele pode dizer "mudo" para silenciar o dispositivo 10 rapidamente durante toda a reunião.

[0059] A resposta à chamada por controle de voz assistido por sensor funciona como segue. Quando o aplicativo que gerencia a chamada recebida recebe uma chamada recebida, o previsor de atividade 18 é instruído para detectar quando a atividade indicativa de elevar o telefone ocorre. Quando o usuário não eleva o dispositivo 10, o previsor de atividade 18 reflete a ocorrência da atividade indicativa, assim o dispositivo 10 pode prever que o usuário vai falar a palavra-chave para atender por voz, e permitir que as palavras chaves para atender por voz sejam reconhecidas mais facilmente. Por exemplo, o tomador de decisão 22 reduz um limiar para confirmar o reconhecimento positivo da palavra-chave para atender por voz, assim a palavra-chave para atender por voz pode ser reconhecida mesmo quando é dita fracamente ou em ambientes ruidosos. Uma vez que o tomador de decisão 22 reflete que a palavra-chave para atender

por voz é reconhecida, o dispositivo 10 pode reagir adequadamente, por exemplo, aceitar, rejeitar, ignorar ou silenciar a chamada. De modo contrário, quando a atividade indicativa de elevar o telefone não é detectada, é improvável que o usuário diga a palavra-chave para atender por voz, assim o dispositivo 10 pode evitar o reconhecimento errôneo da palavra-chave para atender por voz. Por exemplo, o tomador de decisão 22 aumenta o limiar para confirmar o reconhecimento das palavras chaves para atender por voz.

[0060] FIG. 6 demonstra um cenário de fazer uma chamada transmitida. Quando o usuário quer fazer uma chamada transmitida, uma sequência de ação intuitiva e conveniente é: elevar o dispositivo 10 para o lado da orelha, e dizer um nome de quem se quer chamar, por exemplo, "ligar para Mary". Para implementar uma chamada de telefone por controle de voz assistida por sensor, uma lista de atividade correspondente a um aplicativo de comunicação responsável por gerenciar chamada transmitida pode incluir um modelo de atividade que grava a atividade indicativa de elevar o telefone para o lado da orelha, e a lista de palavra-chave correspondente ao aplicativo de comunicação pode incluir as palavras chaves de chamada de voz candidatas relacionadas aos nomes dos possíveis objetos de chamada, por exemplo, nomes listados na agenda e/ou lista de contatos.

[0061] A chamada por controle de voz assistida por sensor funciona como segue. Quando o usuário ativa o aplicativo de comunicação para se preparar para fazer uma chamada transmitida ou quando o dispositivo 10 está executando automaticamente o aplicativo de comunicação como um aplicativo padrão executado quando nenhum outro aplicativo está funcionando, o previsor de atividade 18 é informado para detectar quando a

atividade indicativa de elevar o telefone ocorre. Quando o usuário não eleva o dispositivo 10 para o lado da orelha, o previsor de atividade 18 reflete a ocorrência da atividade indicativa, assim o dispositivo 10 pode prever que o usuário vai falar uma palavra-chave de chamada de voz, e portanto permite que a palavra-chave de chamada de voz seja reconhecida mais facilmente. Por exemplo, o tomador de decisão 22 aumenta a tendência a admitir o reconhecimento positivo da palavra-chave para atender por voz, assim a palavra-chave de chamada de voz pode ser reconhecida mesmo quando é dita fracamente ou em ambientes ruidosos. Uma vez que o tomador de decisão 22 reflete que uma palavra-chave de chamada de voz é reconhecida, o dispositivo 10 faz a chamada de acordo com a palavra-chave de chamada de voz. Por outro lado, quando a atividade indicativa não é detectada, é improvável que o usuário diga uma palavra-chave de chamada de voz, assim o dispositivo 10 pode evitar o reconhecimento errôneo duma palavra-chave de chamada de voz; equivalentemente, o tomador de decisão 22 pode reduzir a tendência a admitir o reconhecimento da palavra-chave para atender por voz.

[0062] FIG. 7 demonstra um cenário de acordar um telefone inativo e/ou destravar um telefone bloqueado. Para economizar energia, smartphones modernos entram em um modo de repouso em que apenas as funções básicas são mantidas, e outras funções (por exemplo, exibição de tela) são desligadas. Para maior comodidade do usuário, é desejado despertar um telefone do modo de repouso simplesmente pela voz do usuário, em vez de tocar a tela ou o botão de empurrar. Assim, o usuário pode despertar o telefone do repouso dizendo uma palavra-chave, ex, "acordar". Para implementar consequentemente um despertar do telefone por controle de voz assistido por sensor, uma lista de atividade

correspondente a um aplicativo de espera responsável pela manipulação do despertar do telefone pode incluir modelo(s) de atividade que grava um estado(s) indicativo no qual o dispositivo 10 é mantido estável, quieto e estando a mão, e a lista de palavra-chave correspondente ao aplicativo de espera pode incluir palavra-chave de despertar por voz candidata como "acordar."

[0063] O despertar do telefone por controle de voz assistido por sensor funciona como segue. Quando o dispositivo 10 vai para um modo de repouso e o aplicativo de espera está funcionando, o previsor de atividade 18 é informado para detectar quando qualquer dos estados indicativos ocorre. Quando o dispositivo 10 não entra em um dos estados indicativos, o previsor de atividade 18 reflete a entrada do estado indicativo, assim o dispositivo 10 pode esperar a palavra-chave de despertar por voz, e portanto permite que a palavra-chave de despertar por voz seja reconhecida mais facilmente. Por exemplo, o tomador de decisão 22 tende a aceitar o reconhecimento positivo da palavra-chave de despertar por voz, assim a palavra-chave de despertar por voz pode ser reconhecida mesmo quando é dita fracamente ou em ambientes ruidosos. Uma vez que o tomador de decisão 22 reflete que a palavra-chave de despertar por voz é reconhecida, o dispositivo 10 pode sair do modo de repouso. Por outro lado, quando nenhum dos estados indicativos é detectado, por exemplo, quando o dispositivo 10 é carregado em uma mochila, é improvável que o usuário diga a palavra-chave de despertar por voz, assim o dispositivo 10 pode evitar o reconhecimento errôneo da palavra-chave de despertar por voz; de modo equivalente, o tomador de decisão 22 tende a rejeitar ou ignorar o reconhecimento da palavra-chave de despertar por voz.

[0064] Apesar de as FIG. 4 a FIG. 7 apenas ilustrarem um número limite de exemplos, deve-se notar que a detecção da palavra-chave em discurso assistido por sensor da invenção pode realmente fornecer uma função de controle de voz natural, conveniente, intuitiva, amigável, precisa e confiável facilmente na vida diária do usuário. Por exemplo, o usuário pode acordar o dispositivo pela voz (Fig. 7) e, em seguida, tirar foto pela voz (Fig. 4), fazer chamadas transmitidas por voz (Figura 6) e/ou atender chamadas recebidas por voz (Fig. 5).

[0065] A lista(s) de atividade e modelo(s) de atividade correspondente em uma base de dados de atividade 20 (FIG. 1) pode ser preestabelecida por um fabricante ou um projetista do dispositivo 10. Além disso, a base de dados de atividade 20 pode ser modificada de acordo com a necessidade individual do usuário. Por exemplo, o sistema 12 pode permitir que o usuário adicione, apague, edite, mova e/ou substitua o modelo de atividade e lista de atividade. Junto com a FIG. 1, favor consultar a FIG. 8 que ilustra um fluxograma 200 de acordo com uma modalidade da invenção. O dispositivo 10 pode adotar o fluxograma 200 para atualizar a base de dados de atividade 20 pela adição de um novo modelo de atividade ou modificação de um modelo de atividade existente por um novo. O fluxograma 200 inclui as etapas a seguir.

[0066] Etapa 202: pelo dispositivo 10, entrar em um modo de treinamento para se preparar para a modificação do usuário quando o usuário quer atualizar manualmente a base de dados de atividade 20. O dispositivo 10 pode, então, pedir ao usuário para especificar um elemento (por exemplo, uma lista de atividade e/ou um modelo de atividade) a ser modificado, e como o elemento vai ser modificado (por exemplo, por adição ou

substituição). Quando o dispositivo 10 entra no modo de treinamento, o dispositivo 10 pode primeiro parar o fluxograma 100 (FIG. 3).

[0067] Etapa 204: pelo dispositivo 10, coletar dados do sensor quando o usuário realiza uma nova atividade destinada a ser adicionada à base de dados de atividade 20 como um novo modelo de atividade, ou destinada a substituir um modelo de atividade existente. Em uma modalidade, o dispositivo 10 pode extrair recursos dos dados do sensor, por exemplo, pelo previsor de atividade 18 do sinal de dados do sensor Ssd.

[0068] Etapa 206: para estabelecer um modelo de atividade estatisticamente confiável, o dispositivo 10 sugere ao usuário repetir a nova atividade várias vezes; cada vez quando o usuário repete a nova atividade, o dispositivo 10 repete a etapa 204. Quando o dispositivo 10 coleta dados do sensor suficientes para construir um novo modelo de atividade confiável, o dispositivo 10 prossegue para a etapa 208. Se os dados coletados não são satisfatórios, o fluxograma 200 repete a etapa 204.

[0069] Etapa 208: pelo dispositivo 10, atualizar a base de dados de atividade, por exemplo, adicionar o novo modelo de atividade ou substituir um modelo de atividade existente por um novo modelo de atividade, de acordo com os dados do sensor coletados.

[0070] Etapa 210: sair do modo de treinamento, então o dispositivo 10 pode restaurar o fluxograma suspenso 100 (etapa 202), ou entrar em outro modo.

[0071] Além da base de dados de atividade 20, a base de dados de palavra-chave 16 também pode ser modificada pelo usuário de acordo com um fluxograma similar ao fluxograma 200.

[0072] Porque a detecção de palavra-chave em discurso é esperada para monitorar a palavra-chave a qualquer momento sem sugestão do usuário, o consumo de energia é ainda considerado, especialmente para dispositivos móveis que dependem da bateria para fonte de energia. A detecção de voz pode ser aplicada para avaliar como o som que entra informativo é, de modo a ativar a detecção de palavra-chave em discurso quando o som que entra parece ser informativo, e de outra forma desativar a detecção de palavra-chave em discurso para economia de energia.

[0073] Junto com a FIG. 1, favor consultar a FIG. 9 que ilustra a detecção de palavra-chave em discurso assistida por sensor do sistema de baixa energia 32 integrado em um dispositivo 30. Similar ao sistema 12 da FIG. 1, o sistema 32 na FIG. 9 inclui um detector de palavra-chave em discurso 44, uma base de dados de palavra-chave 46, um previsor de atividade 48, uma base de dados de atividade 50 e um tomador de decisão 52. Além disso, o sistema 32 inclui um detector de voz 34. O dispositivo 30 inclui um multiplexador 36, um conversor de analógico para digital (ADC) 38 e dois terminais 40a e 40b. Se o dispositivo 30 é equipado com microfone(s) analógico (não mostrado), o microfone(s) analógico pode ser conectado ao terminal 40a, assim um sinal de som analógico (não mostrado) fornecido pelo microfone(s) analógico pode ser retransmitido para o ADC 38 para ser convertido em um sinal de som digital (não mostrado), que é então é então conduzido para se tornar um sinal de som digital Snd pelo multiplexador 36. Por outro lado, se o dispositivo 30 é equipado com microfone(s) digital (não mostrado), o microfone(s) digital pode ser conectado ao terminal 40b, assim um sinal de som digital (não mostrado) fornecido pelo microfone(s) digital pode ser conduzido para se

tornar o sinal Snd pelo multiplexador 36. O sistema 32 também recebe um sinal de dados do sensor Ssd fornecido pelo sensor(s) do dispositivo 30.

[0074] O funcionamento do detector de palavra-chave em discurso 44, o previsor de atividade 48 e o tomador de decisão 52 são similares àquele do detector de palavra-chave em discurso 14, o previsor de atividade 18 e o tomador de decisão 22 (FIG. 1). Com o funcionamento do detector de palavra-chave em discurso 44, o previsor de atividade 48 e o tomador de decisão 52 assim como as bases de dados 46 e 50 podem implementar a detecção de palavra-chave em discurso assistida por sensor da invenção pelo seguinte fluxograma 100 (FIG. 3).

[0075] A base de dados de palavra-chave 46 é acoplada ao detector de palavra-chave em discurso 44, e inclui um número N2 de listas de palavra-chave KL[1] a KL[N2]. Cada lista de palavra-chave KL[j] (para j = 1 a N2) é associada com um aplicativo alvo app[j] incluindo uma quantidade P{j} de palavras chaves candidatas kw[j,1] a kw[j,P{j}].

[0076] A base de dados de atividade 50 é acoplada ao previsor de atividade 48, e inclui um número N1 de listas de atividade AL[1] a AL[N1]; cada lista de atividade AL[i] (para i = 1 a N1) é associada com um aplicativo alvo app[i] incluindo uma quantidade Q{i} de modelos de atividade at[i,1] a at[i,Q{i}].

[0077] O detector de palavra-chave em discurso 44 obtém uma lista de palavra-chave selecionada KL[jx] das listas de palavra-chave KL[1] a KL[N2] da base de dados de palavra-chave 46, sendo que o aplicativo alvo app[jx] da lista de palavra-chave selecionada KL[jx] corresponde a um aplicativo atualmente em operação do dispositivo 30. Em resposta ao sinal Snd, o detector

de palavra-chave em discurso 44 fornece um resultado de detecção de palavra-chave preliminar Skw de acordo com os lista de palavra-chave selecionada KL[jx].

[0078] O previsor de atividade 48 obtém a lista de atividade selecionada AL[ix] das listas de atividade AL[1] a AL[N1] da base de dados de atividade 50, sendo que o aplicativo alvo app[ix] da lista de atividade selecionada AL[ix] corresponde ao aplicativo atualmente em execução do dispositivo 30. Com base na lista de atividade selecionada AL[ix], o previsor de atividade 48 pode processar o sinal Ssd para fornecer um resultado de previsão de atividade Sap indicando uma probabilidade de se um usuário está prestes a emitir a palavra-chave por voz.

[0079] O tomador de decisão 52 é acoplado ao previsor de atividade 48 e o detector de palavra-chave em discurso 44, para processar os resultados Sap e Skw para fornecer um segundo resultado de detecção de palavra-chave Sdm, assim o dispositivo 30 pode reagir de acordo com o resultado Sdm.

[0080] O detector de voz 34 é acoplado ao detector de palavra-chave em discurso 44, e é capaz de avaliar a informatividade com base em, por exemplo, SNR, do sinal de som Snd, e dessa forma determinar quando ativar o detector de palavra-chave em discurso 44. Por exemplo, se a informatividade do sinal Snd for baixa (por exemplo, abaixo de um limiar de informatividade, não mostrado), o detector de voz 34 pode desabilitar (inativar) o detector de palavra-chave em discurso 44, por exemplo, manter o detector de palavra-chave em discurso 44 em um modo de baixa energia (ocioso). Por outro lado, se a informatividade do sinal Snd é alta (por exemplo, maior do que o limiar de informatividade), o detector de voz 34 pode habilitar (ativar) o detector de palavra-chave em discurso 44, por exemplo,

despertar o detector de palavra-chave em discurso 44 para funcionar um modo (totalmente funcional) normal. Como mostrado na FIG. 9, o detector de voz 34 fornece um sinal Svd para seletivamente ativar o detector de palavra-chave em discurso 44.

[0081] Quando o detector de palavra-chave em discurso 44 é desativado, o previsor de atividade 48 e o tomador de decisão 52 são de preferência desabilitados assim como as bases de dados 46 e 50. Quando o detector de palavra-chave em discurso 44 é ativado, o previsor de atividade 48 e o tomador de decisão 52 (assim como as bases de dados 46 e 50) são habilitados a cooperar com o detector de palavra-chave em discurso 44 para a detecção de palavra-chave em discurso assistida por sensor. Alternativamente, o previsor de atividade 48 e o tomador de decisão 52 também recebem o sinal Svd a ser habilitado ou desabilitado.

[0082] Favor consultar a FIG. 10 que ilustra uma modalidade do detector de voz 34. O detector de voz 34 inclui um estimador de discurso 56a, um estimador de ruído 56b e um comparador 58 acoplado ao estimador de discurso 56a e o estimador de ruído 56b. O estimador de discurso 56a é capaz de fornecer uma sequência $S[.]$ em resposta ao volume do sinal Snd; já que cada amostra do sinal Snd é denotada por um valor $x[n]$, volume do sinal Snd pode ser denotado por um valor absoluto $|x[n]|$. O estimador de ruído 56b é capaz de fornecer uma sequência $N[.]$ também em resposta ao volume do sinal Snd.

[0083] Como mostrado na equação eq1 da FIG. 10, o estimador de discurso 56a é capaz de computar uma amostra atual $S[n]$ da sequência $S[.]$ como uma soma ponderada de um volume atual $|x[n]|$ do sinal Snd e uma amostra anterior $S[n-1]$ da sequência

$S[.]$, isto é, computar a amostra atual $S[n]$ da sequência $S[.]$ pela soma de um produto $a0*S[n-1]$ e um produto $(1-a0)*|x[n]|$, em que o produto $a0*S[n-1]$ é um resultado da multiplicação da amostra anterior $S[n-1]$ com um peso $a0$, e o produto $(1-a0)*|x[n]|$ é um resultado da multiplicação do volume atual $|x[n]|$ por um peso complementar $(1-a0)$. Em uma modalidade, o peso $a0$ pode ser maior que 0 e menor que 1.

[0084] Como mostrado na equação eq2 da FIG. 10, o estimador de ruído 56b é capaz de computar uma amostra atual $N[n]$ da sequência $N[.]$ como uma soma ponderada do valor atual $|x[n]|$ e uma amostra anterior $N[n-1]$ da sequência $N[.]$, isto é, computando a amostra atual $N[n]$ da sequência $N[.]$ pela soma de um produto $b0*N[n-1]$ e um produto $(1-b0)*|x[n]|$, em que o produto $b0*S[n-1]$ é um resultado da multiplicação da amostra anterior $N[n-1]$ com um peso $b0$, e o produto $(1-b0)*|x[n]|$ é um resultado da multiplicação do volume atual $|x[n]|$ por um peso complementar $(1-b0)$. Em uma modalidade, o peso $b0$ pode ser maior que 0 e menor que 1.

[0085] Em uma modalidade, o peso $a0$ é menor que o peso $b0$. Assim, a sequência $S[.]$ tende refletir o volume atual do som que entra, e a sequência $N[.]$ tende a refletir o volume médio passado do som que entra. Assim, a sequência $S[.]$ indica o sinal informativo, por exemplo, voz, contida no sinal Snd enquanto a sequência $N[.]$ indica ruído de fundo no sinal Snd. O comparador 58 compara $S[.]$ e $N[.]$ para indicar a informatividade do sinal Snd, e assim fornecer o sinal Svd para controlar o detector de palavra-chave em discurso 44. Por exemplo, se uma diferença absoluta $|S[n]-N[n]|$ é maior do que um limiar de informatividade (não mostrado), então o comparador 58 do detector de voz 34 reflete a alta informatividade no sinal Svd para ativar o detector de

palavra-chave em discurso 44. De modo contrário, se a diferença absoluta $|S[n] - N[n]|$ é menor que o limiar de informatividade, então o comparador 58 do detector de voz 34 reflete uma baixa informatividade no sinal Svd para desabilitar o detector de palavra-chave em discurso 44 porque a execução da detecção de palavra-chave em discurso em som ruidoso só leva a erro.

[0086] Junto com a FIG. 9, favor consultar a FIG. 11 que ilustra um fluxograma 300 de detecção seletiva de palavra-chave em discurso de acordo com uma modalidade da invenção. Com o detector de voz 34, o sistema 32 (FIG. 9) pode adotar o fluxograma 300 para economia de energia sem comprometer a função esperada da detecção de palavra-chave em discurso. As maiores etapas do fluxograma 300 podem ser descritas como segue.

[0087] Etapa 302: pelo detector de voz 34, detectar a voz no som; por exemplo, avaliar a informatividade do sinal Snd.

[0088] Etapa 304: se a voz é detectada (informatividade é alta), prosseguir para a etapa 306, caso contrário prosseguir para a etapa 308.

[0089] Etapa 306: realizar a detecção de palavra-chave em discurso, por exemplo, a detecção de palavra-chave em discurso assistida por sensor do fluxograma 100 (FIG. 3). Embora o detector de voz 34 seja utilizado junto com detecção de palavra-chave em discurso assistida por sensor do detector de palavra-chave em discurso 44, o previsor de atividade 48 e o tomador de decisão 52 como mostrado na FIG. 9, o detector de voz 34 também pode ser usado junto com uma detecção de palavra-chave em discurso mais simples que é implementado somente por um detector de palavra-chave em discurso preliminar (por exemplo, o detector de palavra-chave em discurso 44) sem

assistência dos dados do sensor. Tal detecção de palavra-chave em discurso mais simples também pode adotar o fluxograma 300, e a detecção de palavra-chave do discurso habilitada na etapa 306 é implementada pela operação do detector de palavra-chave em discurso preliminar. Após a etapa 306, o fluxograma 300 pode repetir a etapa 302.

[0090] Etapa 308: desabilitar a detecção de palavra-chave em discurso, por exemplo, parar o funcionamento do detector de palavra-chave em discurso 44 (FIG. 9). O funcionamento do previsor de atividade 48 e o tomador de decisão 52, assim como a base de dados de palavra-chave 46 e a base de dados de atividade 50, também pode ser interrompido. Após a etapa 308, o fluxograma 300 pode repetir a etapa 302.

[0091] Em uma modalidade, o detector de voz 34 na FIG. 10 pode ser implementado por um hardware dedicado, por exemplo, circuito. O detector de palavra-chave em discurso 44, o previsor de atividade 48 e/ou o tomador de decisão 52 (FIG. 9) também pode ser implementado pelo hardware dedicado, ou por um processador de sinal digital (DSP). O DSP pode executar software e/ou firmware para implementar as funções do detector de palavra-chave em discurso 44, o previsor de atividade 48 e/ou o tomador de decisão 52. A base de dados de palavra-chave 46 e a base de dados de atividade 50 podem ser implementadas pela memória não volátil programável. Um dispositivo moderno pode incluir uma CPU e um DSP. A CPU é responsável pela execução do OS e aplicativos, e o DSP é responsável pelo processamento multimídia (por exemplo, codificação e / ou decodificação de áudio e/ou vídeo). Sob tal arquitetura, os componentes do sistema de detecção de palavra-chave em discurso assistida por sensor podem ser implementados no DSP, de modo que a detecção de

palavra-chave em discurso assistida por sensor possa funcionar sem envolvimento da CPU, e dessa forma reduzir o consumo de energia da detecção de palavra-chave em discurso assistida por sensor. Alternativamente, só o detector de voz 34 pelo hardware dedicado é arranjado para funcionar na maioria do tempo para seletivamente ativar o circuito necessário. Por exemplo, quando um smartphone aumenta para uma sensibilidade de detecção de voz como divulgado pela invenção, o smartphone pode solicitar ao usuário por um LED com cor diferente, por exemplo, laranja, que complete uma operação desejada pelo controle de voz sob a tela escura. Alternativamente, a tela pode mostrar um logotipo indicativo dentro de uma área parcial de uma tela cheia para solicitar ao usuário. Alternativamente, apenas um sensor giroscópico é arranjado para sempre ou periodicamente funcionar para seletivamente ativar circuito necessário uma vez que o giroscópico consome muito pouca energia. De acordo com uma pluralidade de operações definidas pela base de dados de atividade e a base de dados de palavra-chave, o smartphone pode operar sob a tela escura ou não requerer a iluminação completa da tela. Assim, o consumo de energia pode ser insignificante ou até reduzido ao realizar o conceito divulgado pela invenção.

[0092] Em suma, a invenção aproveita dados do sensor fornecido pelo sensor(s) do dispositivo para obter a informação adicional para melhorar a detecção de palavra-chave em discurso, de modo a implementar um controle de voz mais conveniente, amigável, confiável e preciso. Pessoas versadas na técnica podem fazer modificações possíveis, sem se afastar do escopo revelado pela invenção. Por exemplo, o controle de toque da tela escura pode ser incorporado com a invenção para aumentar ainda mais a conveniência do dispositivo de controle.

[0093] Enquanto a invenção foi descrita em termos do que é atualmente considerado como as modalidades mais práticas e preferidas, deve ser entendido que a invenção não precisa ser limitada à modalidade descrita. Pelo contrário, pretende-se cobrir várias modificações e arranjos semelhantes incluídos dentro do espírito e escopo das reivindicações anexas, que devem ser reconhecidas com a interpretação mais ampla de modo a abranger todas essas modificações e estruturas semelhantes.

REIVINDICAÇÕES

1. Sistema (10) para detecção de palavra-chave em discurso de um dispositivo, caracterizado por compreender:

um previsor de atividade (18) capaz de obter dados do sensor (Ssd) fornecidos por uma pluralidade de sensores e capaz de processar os dados do sensor para fornecer um resultado de previsão de atividade (Sap) indicando uma ocorrência de atividades indicativas que refletem se um usuário está prestes a emitir uma palavra-chave por voz;

um detector de palavra-chave (14) em discurso para gerar uma mais provável palavra-chave como resultado de detecção de palavra-chave preliminar (Skw); e

um tomador de decisão (22) acoplado ao previsor de atividade e o detector de palavra-chave em discurso, capaz de processar o resultado de previsão de atividade e o resultado de detecção de palavra-chave preliminar para fornecer um resultado de detecção de palavra-chave (Sdm) medindo o quanto provável a mais provável palavra-chave é realmente dita, e capaz de diminuir uma sensibilidade da identificação da mais provável palavra-chave em que as atividades indicativas não são detectadas, para então evitar erro de alarme falso.

2. Sistema, de acordo com a reivindicação 1, caracterizado por adicionalmente compreender:

uma base de dados de atividade (20) compreendendo uma pluralidade de listas de atividade (AL[.]); cada lista de atividade sendo associada com um aplicativo alvo (app[.]), e compreendendo uma pluralidade de modelos de atividade (at[.,.]),

em que o previsor de atividade é capaz de:

obter uma selecionada dentre a lista de atividade, em que o aplicativo alvo associado com a lista de atividade

selecionada corresponde a um aplicativo atualmente em operação do dispositivo; e comparar os dados do sensor com cada modelo da atividade da lista de atividade selecionada para fornecer o resultado de previsão de atividade.

3. Sistema, de acordo com a reivindicação 1 ou 2 caracterizado por adicionalmente compreender:

uma base de dados de atividade compreendendo uma pluralidade de listas de atividade; cada lista de atividade sendo associada com um aplicativo alvo, e compreendendo uma pluralidade de modelos de atividade,

em que o previsor de atividade é capaz de:

obter uma selecionada dentre as listas de atividade, em que o aplicativo alvo associado com a lista de atividade selecionada corresponde a um aplicativo atualmente em operação do dispositivo;

fornecer dados extraídos pelos recursos de extração dos dados do sensor; e

comparar os dados extraídos com cada modelo da atividade da lista de atividade selecionada para fornecer o resultado de previsão de atividade.

4. Sistema, de acordo com qualquer uma das reivindicações 1 a 3, caracterizado por adicionalmente compreender:

uma base de dados de palavra-chave (16) compreendendo uma pluralidade de listas de palavra-chave (KL[.]); cada lista de palavra-chave sendo associada com um aplicativo alvo (app[.]), e compreendendo uma pluralidade de palavras chaves candidatas (kw[.,.]),

em que o detector de palavra-chave em discurso é capaz de:

obter uma selecionada dentre as listas de palavra-chave, em que o aplicativo alvo associado com uma lista de palavra-chave selecionada corresponde a um aplicativo atualmente em operação do dispositivo; e

comparar o som que entra com cada palavra-chave candidata da lista de palavra-chave para fornecer o resultado de detecção de palavra-chave preliminar.

5. Sistema, de acordo com qualquer uma das reivindicações 1 a 4, caracterizado por adicionalmente compreender:

um detector de voz (34) capaz de avaliar informatividade do som que entra para seletivamente ativar o detector de palavra-chave em discurso;

um primeiro estimador (56a) capaz de computar uma amostra atual ($S[n]$) de uma primeira sequência pela soma uma primeira parcela e uma segunda parcela, a primeira parcela multiplicando uma amostra precedente ($S[n-1]$) da primeira sequência com um primeiro peso (a_0), e a segunda parcela multiplicando um volume atual ($|x[n]|$) de som entrando com um primeiro peso complementar ($1-a_0$);

um segundo estimador (56b) capaz de computar uma amostra atual ($N[n]$) de uma segunda sequência pela soma de uma terceira parcela e quarta parcela, a terceira parcela multiplicando a amostra precedente ($N[n-1]$) da segunda sequência com um segundo peso (b_0), e a quarta parcela multiplicando um volume atual ($|x[n]|$) de som entrando com um segundo peso complementar ($1-b_0$); e

um comparador (58) acoplado ao primeiro estimador e o segundo estimador comparando a primeira sequência e a segunda sequência para indicar a informatividade do som que entra.

6. Sistema, de acordo com a reivindicação 5, caracterizado por o primeiro peso ser menor do que o segundo peso.

7. Sistema, de acordo com qualquer uma das reivindicações 1 a 6, caracterizado por o tomador de decisão processar a previsão de atividade resultante e a detecção de palavra-chave preliminar resulta no fornecimento da detecção da palavra-chave resultante pela aplicação de uma das seguintes: um algoritmo de fusão de probabilidade, uma máquina de aprendizado de algoritmo e cálculo de uma combinação linear do resultado de previsão de atividade e do resultado de detecção de palavra-chave preliminar como o detector de palavra-chave resultante.

8. Sistema, de acordo com qualquer uma das reivindicações 1 a 7, caracterizado por os sensores incluirem pelo menos um dos seguintes: um acelerômetro, um giroscópio, um magnetômetro, um barômetro um sensor de proximidade, sensor de luz, uma tela sensível ao toque, um receptor de um sistema de posicionamento, um receptor sem fio e uma câmera.

9. Método para detecção de palavra-chave em discurso de um dispositivo, compreendendo obter dados do sensor fornecidos por uma pluralidade de sensores;

caracterizado por, de acordo com os dados do sensor, computar um resultado de previsão de atividade (Sap) indicando uma ocorrência de atividades indicativas que reflete se um usuário está prestes a emitir a palavra-chave por voz; e

computar um resultado de detecção de palavra-chave (Sdm) de acordo com o resultado de previsão de atividade e um resultado de detecção de palavra-chave preliminar (Skw) da detecção de palavra-chave em discurso para medição do quão

provável a mais provável palavra-chave é realmente dita, e diminuir uma sensibilidade da identificação da mais provável palavra-chave em que as atividades indicativas não são detectadas, para então evitar erro de alarme falso.

10. Método, de acordo com a reivindicação 9 caracterizado por adicionalmente compreender:

fornecer uma base de dados de atividade (20) em que compreende uma pluralidade de listas de atividade (AL[.]); cada lista de atividade sendo associada com um aplicativo alvo (app[.]), e compreendendo uma pluralidade de modelos de atividade (at[.,.]);

identificar o aplicativo atualmente em execução do dispositivo;

selecionar uma das listas de atividades pela comparação do aplicativo atualmente em execução com cada aplicativo alvo de cada lista de atividade; e

comparar os dados do sensor com cada modelo da atividade da lista de atividade selecionada para fornecer o resultado de previsão de atividade.

11. Método, de acordo com a reivindicação 9 ou 10 caracterizado por adicionalmente compreender:

fornecer uma base de dados de atividade que compreende uma pluralidade de listas de atividade;

cada lista de atividade sendo associada com um aplicativo alvo e compreendendo uma pluralidade de modelos de atividade;

identificar o aplicativo atualmente em execução do dispositivo;

pela comparação do aplicativo atualmente em execução com cada aplicativo alvo de cada lista de atividade, selecionando

uma das listas de atividade;

fornecer dados extraídos pelos recursos de extração dos dados do sensor;

comparar os dados extraídos com cada modelo da atividade da lista de atividade selecionada e consequentemente fornecer um resultado correspondente de atividade para ser incluído no resultado de previsão de atividade.

12. Método, de acordo com qualquer uma das reivindicações 9 a 11, caracterizado por adicionalmente compreender:

fornecer uma base de dados de palavra-chave (16) que compreende uma pluralidade de listas de palavra-chave (KL[.]); cada lista de palavra-chave sendo associada com um aplicativo alvo (app[.]) e compreendendo uma pluralidade de palavras-chave candidatas (kw[.,]);

identificar o aplicativo atualmente em execução do dispositivo;

selecionar uma das listas de palavra-chave pela comparação do aplicativo atualmente em execução com cada aplicativo alvo de cada lista de palavra-chave; e

fornecer o resultado de detecção de palavra-chave preliminar pela comparação do som que entra com cada palavra-chave candidata da lista de palavra-chave selecionada.

13. Método, de acordo com qualquer uma das reivindicações 9 a 12, caracterizado por adicionalmente compreender:

executar uma detecção de voz pela avaliação da informatividade do som entrando para seletivamente ativar a detecção de palavra-chave do discurso, em que a etapa de executar a detecção de voz compreende:

computar uma amostra atual ($S[n]$) de uma primeira sequência pela soma de um primeiro produto e um segundo produto, o primeiro produto multiplicando uma amostra anterior ($S[n-1]$) da primeira sequência por um primeiro peso (a_0), e o segundo produto multiplicando um volume atual ($|x[n]|$) do som que entra por um primeiro peso de complementaridade ($1-a_0$);

computar uma amostra atual ($N[n]$) de uma segunda sequência pela soma de um terceiro produto e um quarto produto, o terceiro produto multiplicando uma amostra anterior ($N[n-1]$) da segunda sequência por um segundo peso (b_0), e o quarto produto multiplicando um volume atual do som que entra ($|x[n]|$) por um segundo peso de complementaridade ($1-b_0$); e

comparar a amostra atual da primeira sequência e a amostra atual da segunda sequência para indicar a informatividade do som que entra.

14. Método, de acordo com qualquer uma das reivindicações 9 a 13, caracterizado por computar a detecção de palavra-chave do resultado compreender uma das seguintes:

adotar um processo de probabilidade de fusão de algoritmo para processar a atividade resultante de previsão e detectar a resultante de palavra-chave preliminar;

adotar uma máquina de aprendizado de algoritmo para processar a atividade resultante de previsão e a detecção resultante de palavra-chave preliminar, e

calcular uma combinação linear da atividade resultante de previsão e o resultado de detecção de palavra-chave preliminar como o resultado de detecção de palavra-chave.

15. Método, de acordo com qualquer uma das reivindicações 9 a 14, caracterizado por os sensores incluirem pelo menos um dentre os seguintes: um acelerômetro, um

giroscópio, um magnetômetro, um barômetro um sensor de proximidade, sensor de luz, uma tela sensível ao toque, um receptor de um sistema de posicionamento, um receptor sem fio e uma câmera.

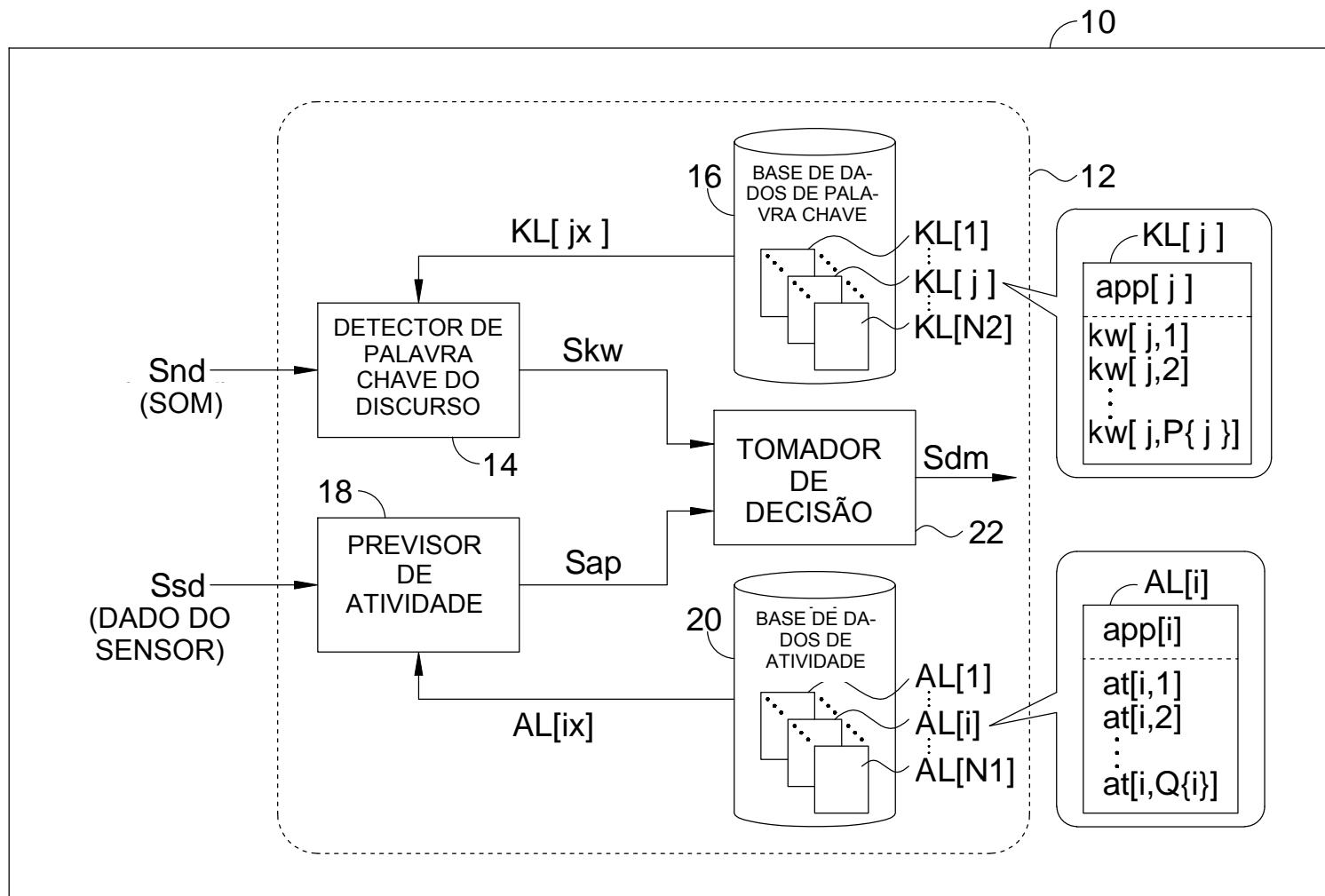


FIG. 1

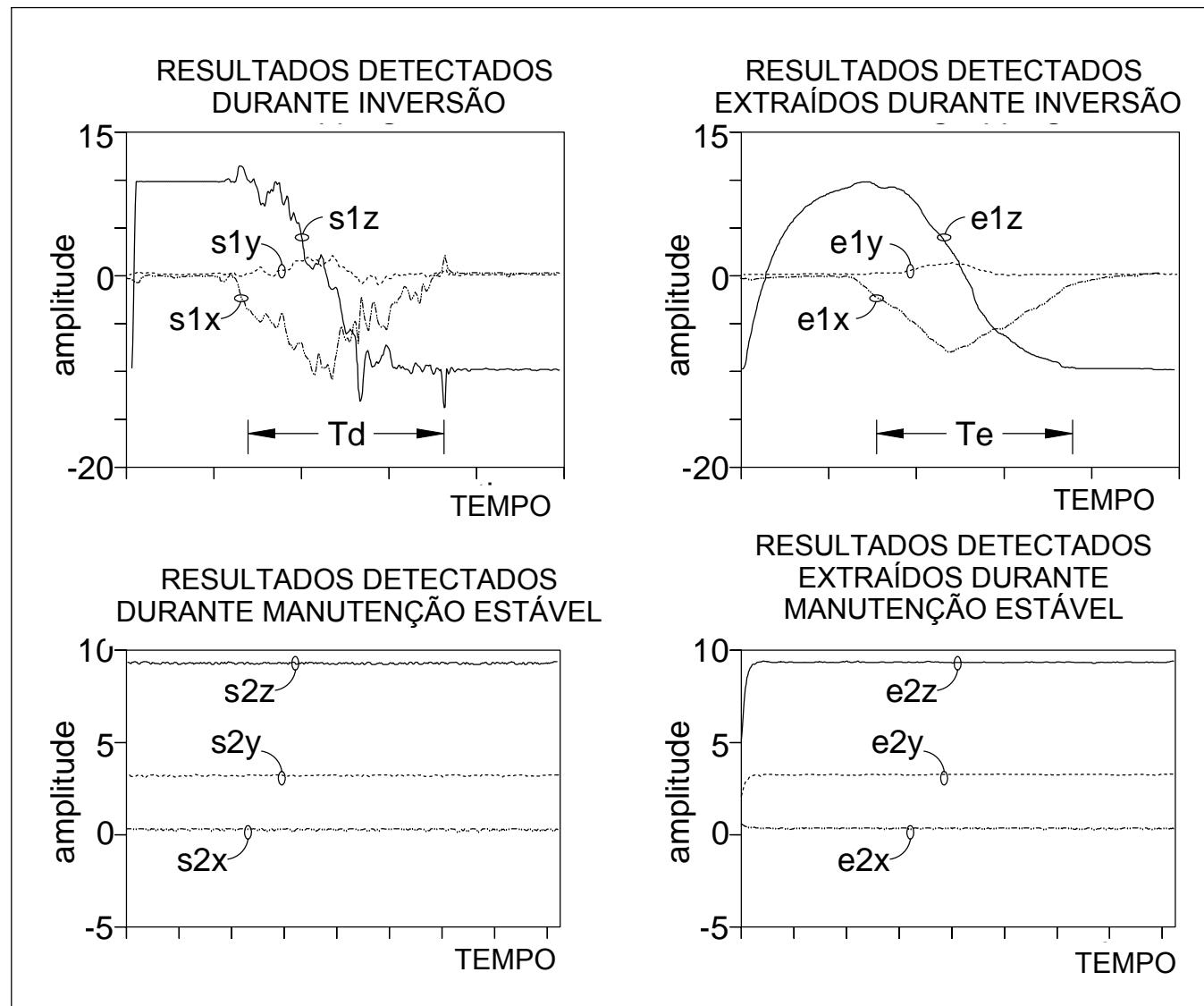


FIG. 2

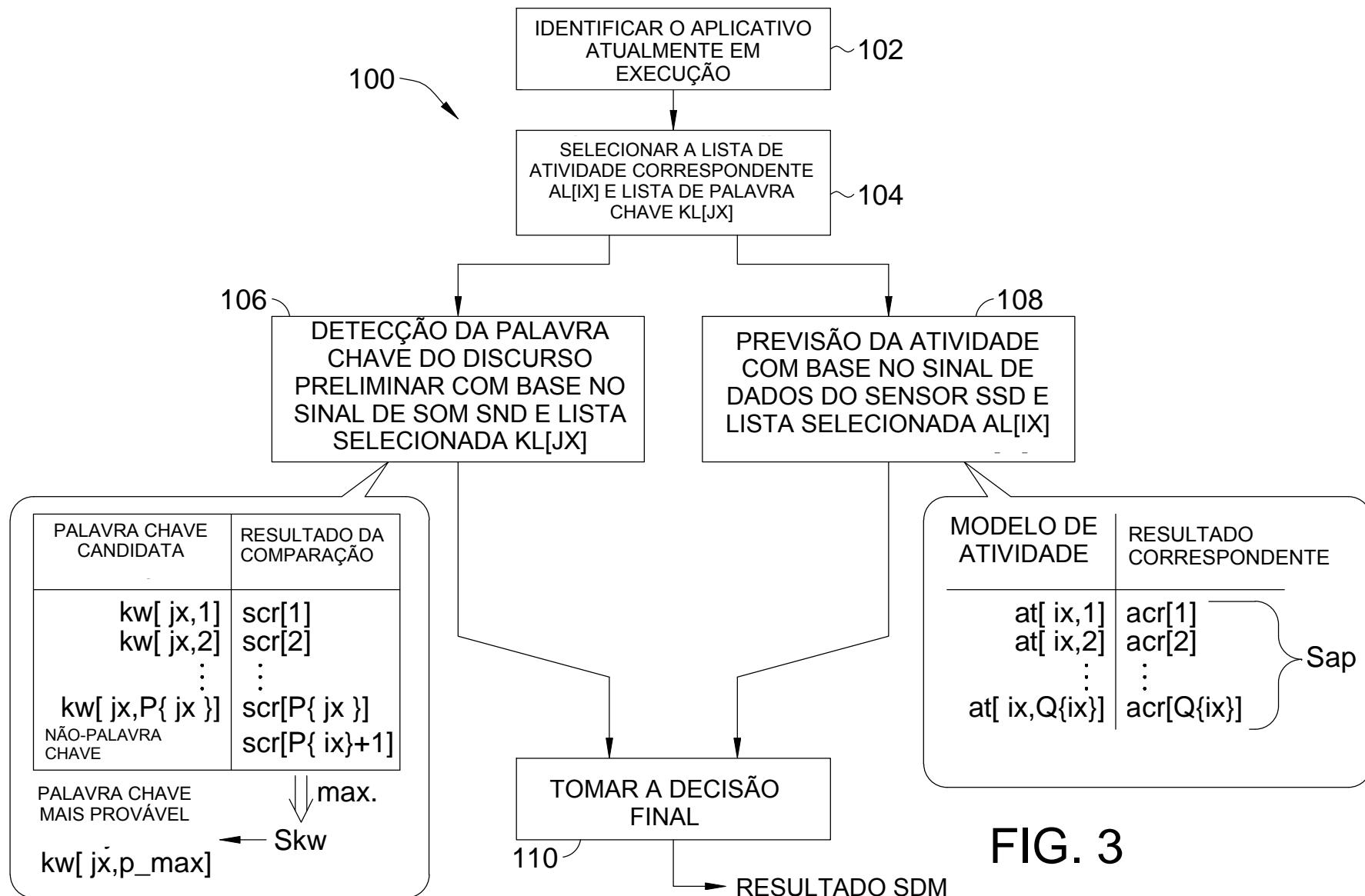


FIG. 3

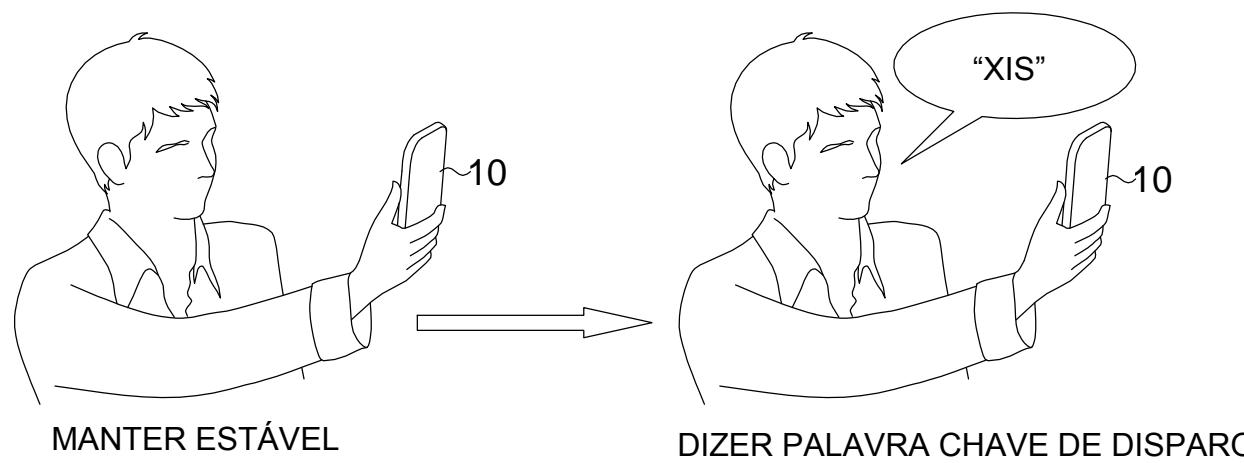


FIG. 4

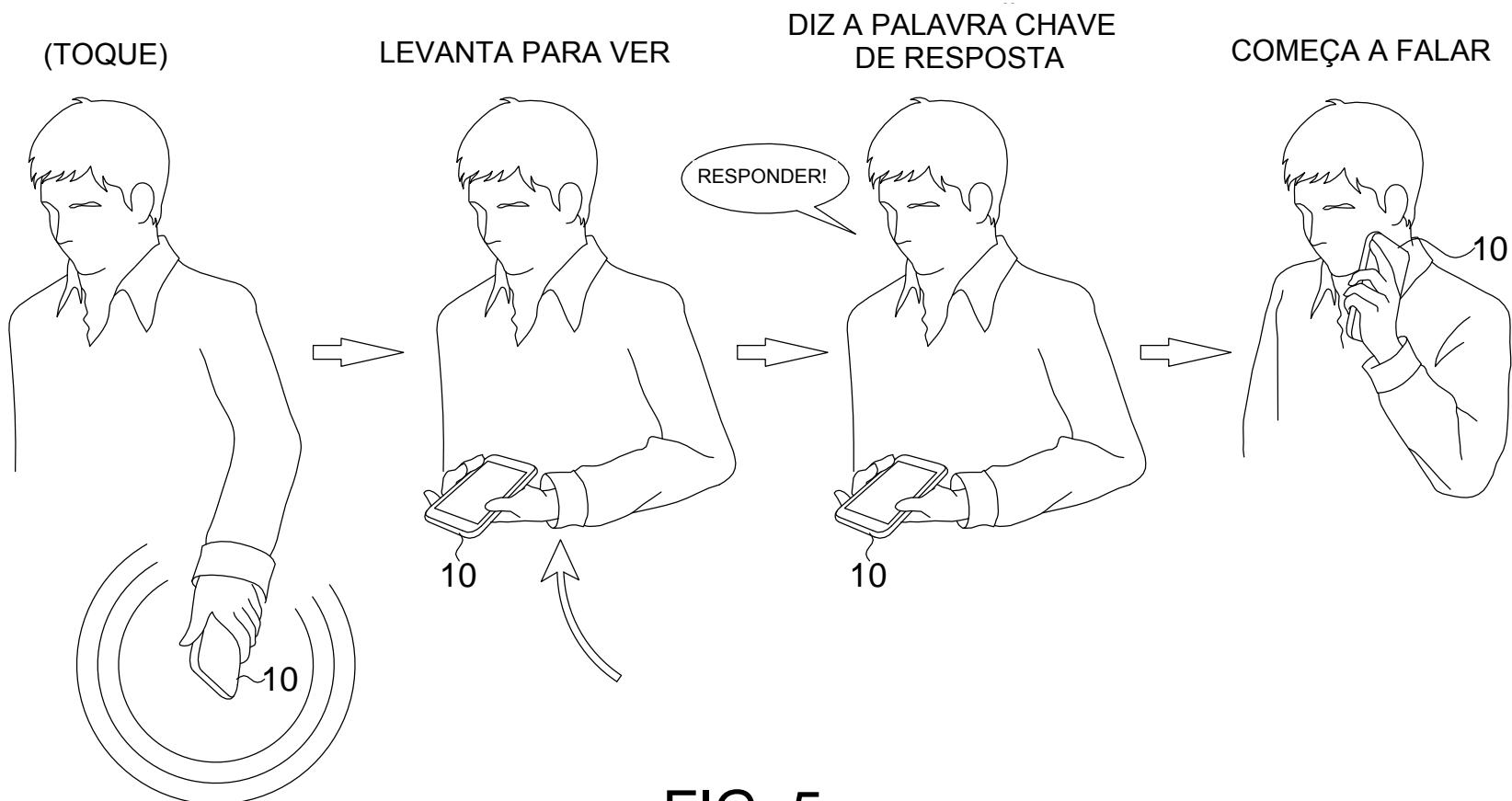
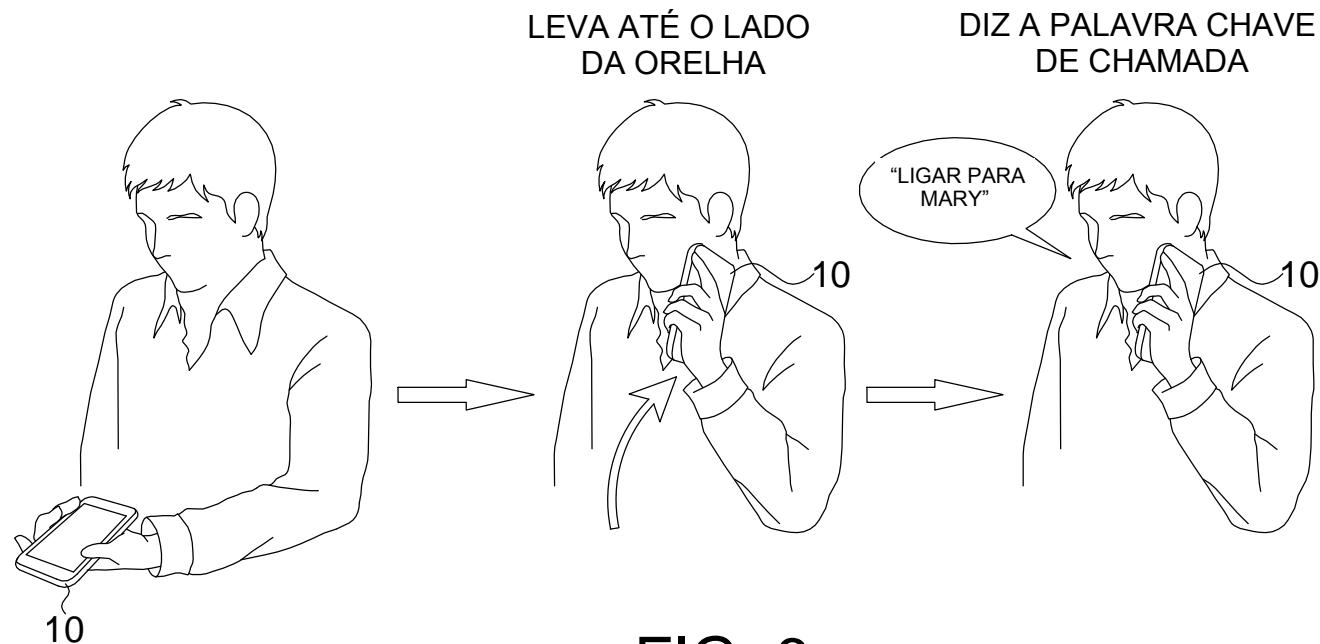


FIG. 5



DIZ PALAVRA CHAVE PARA DESPERTAR

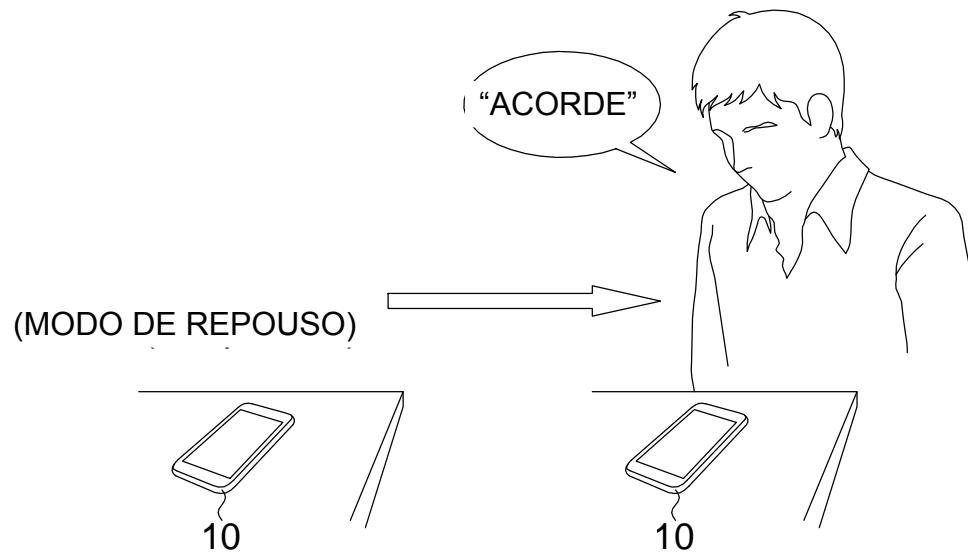


FIG. 7

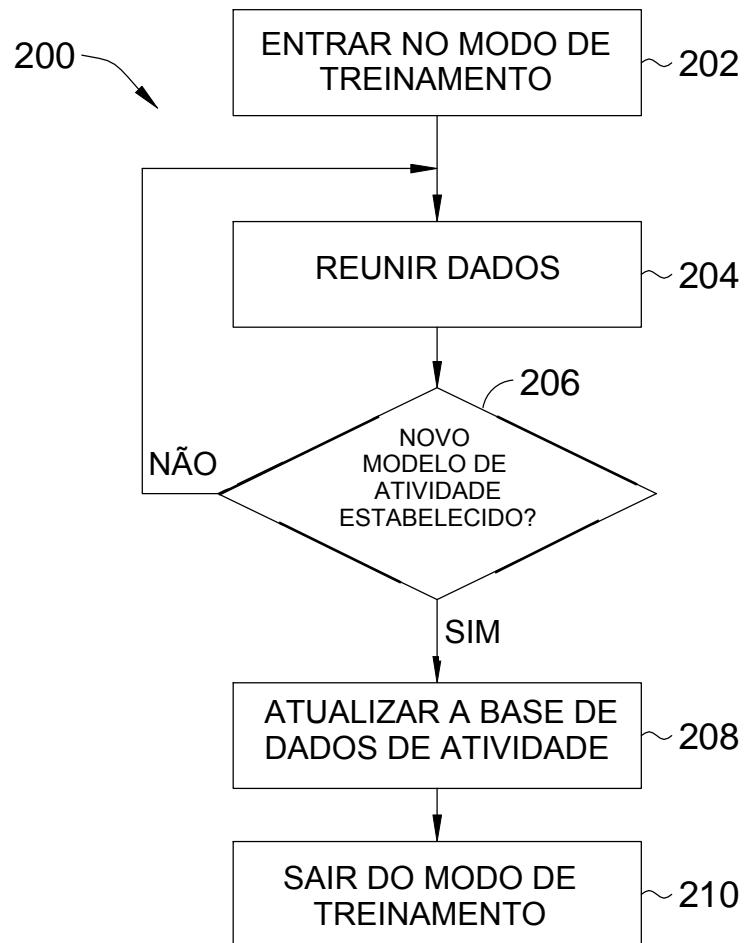


FIG. 8

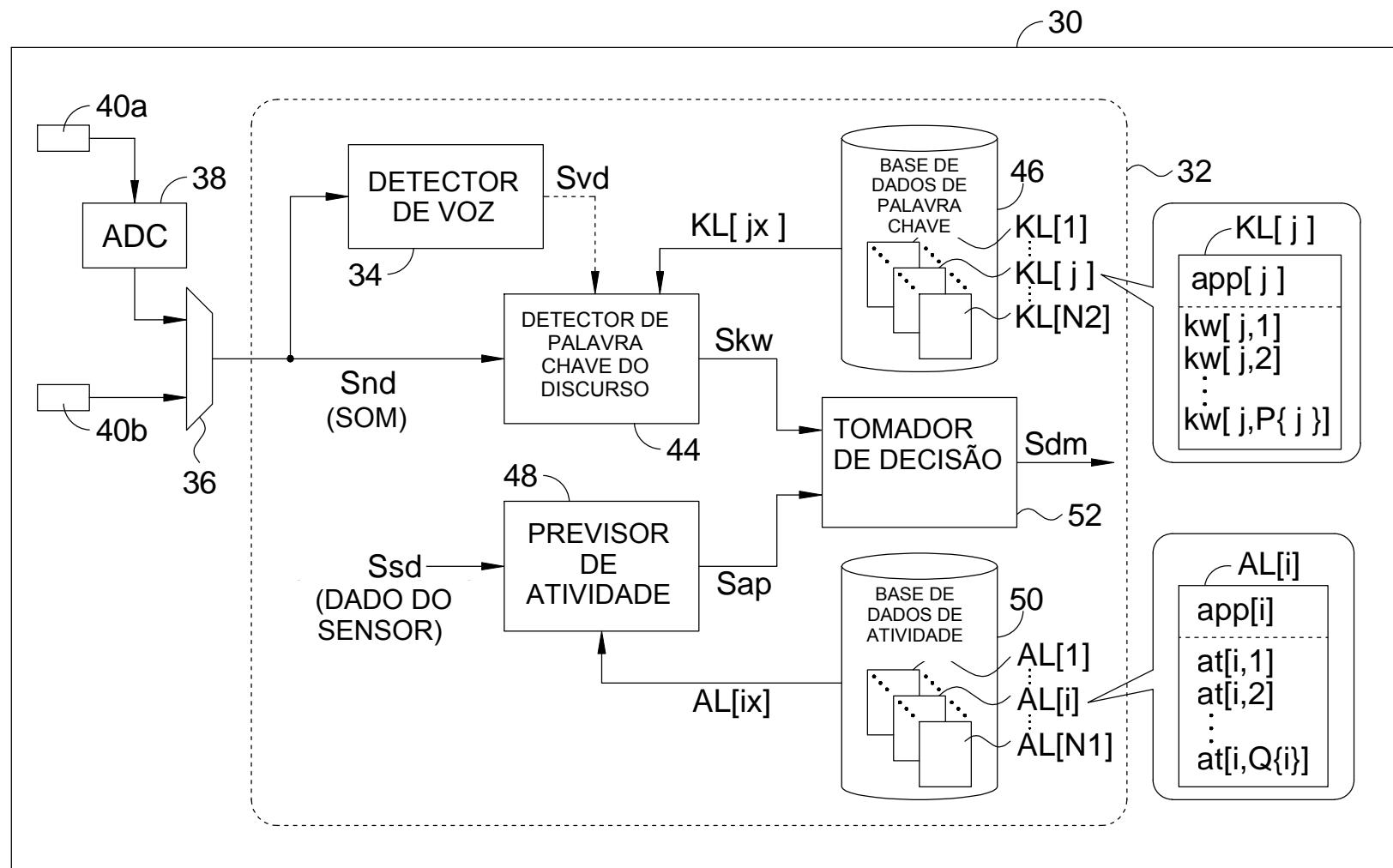


FIG. 9

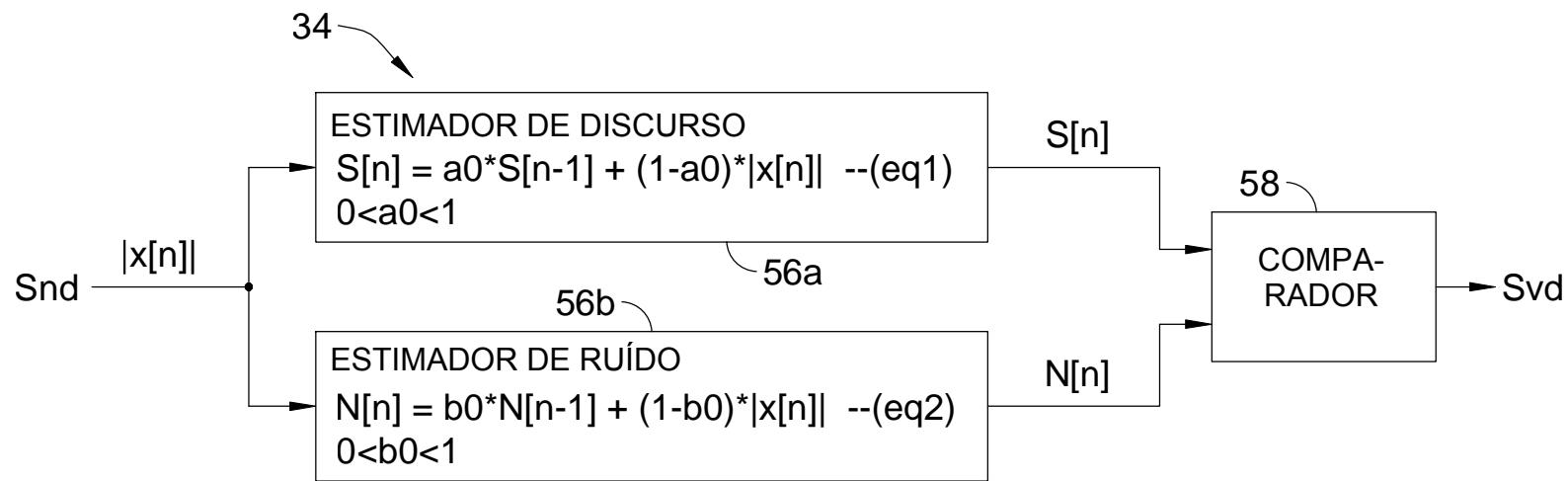


FIG. 10

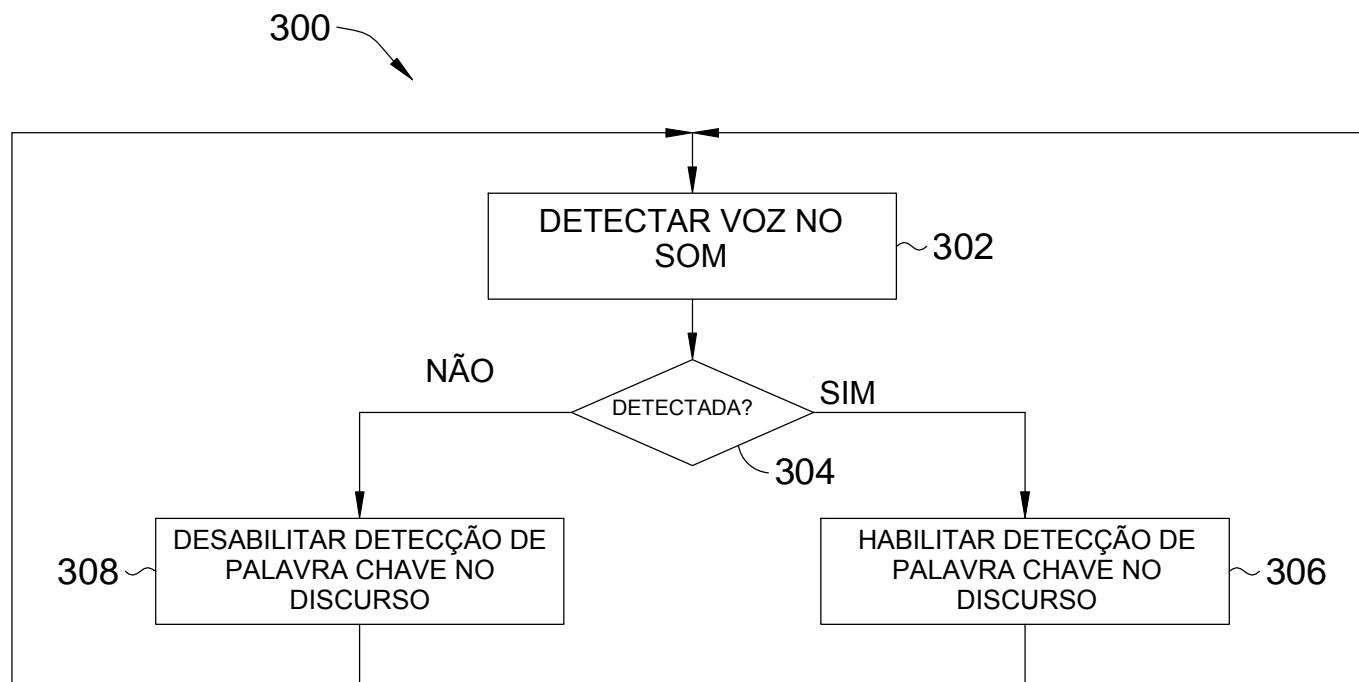


FIG. 11