

República Federativa do Brasil
Ministério do Desenvolvimento, Indústria
e do Comércio Exterior
Instituto Nacional da Propriedade Industrial.

(21) **PI 0714140-8 A2**

(22) Data de Depósito: 15/05/2007

(43) Data da Publicação: 26/12/2012
(RPI 2190)



(51) *Int.Cl.:*

G10L 15/22

G10L 15/00

G10L 15/06

H04M 1/67

(54) **Título:** DETECÇÃO DE UMA SECRETÁRIA ELETRÔNICA UTILIZANDO RECONHECIMENTO DE FALA

(30) **Prioridade Unionista:** 12/07/2006 US 11485011

(73) **Titular(es):** Microsoft Corporation

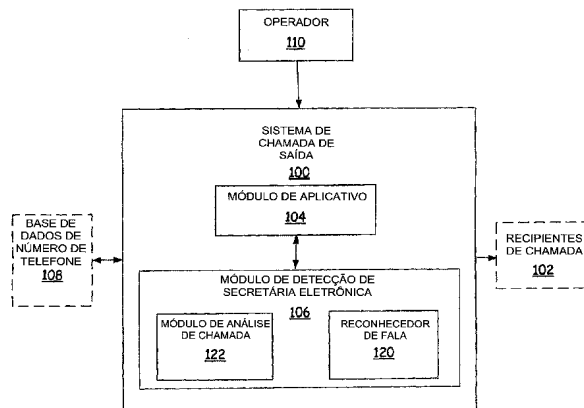
(72) **Inventor(es):** Alejandro Acero, Craig M. Fischer, Gong Yu, Ye-Yi Wang, Yu-Cheng Ju

(74) **Procurador(es):** Ricardo Pinho

(86) **Pedido Internacional:** PCT US2007011567 de 15/05/2007

(87) **Publicação Internacional:** WO 2008/008117 de 17/01/2008

(57) **Resumo:** DETECÇÃO DE UMA SECRETÁRIA ELETRÔNICA UTILIZANDO RECONHECIMENTO DE FALA. Um módulo de detecção de secretária eletrônica é utilizado para determinar se um recipiente de uma chamada é uma pessoa ou uma secretária eletrônica. O módulo de detecção de secretária eletrônica inclui um reconhecedor de fala e um módulo de análise de chamada. O reconhecedor de fala recebe uma resposta audível do recipiente da chamada para uma chamada. O reconhecedor de fala processa a resposta audível e fornece uma saída indicativa da fala reconhecida. O módulo de análise de chamada processa a saída do reconhecedor de fala para gerar uma saída indicativa de se o recipiente de chamada é uma pessoa ou uma secretária eletrônica.



"DETECÇÃO DE UMA SECRETÁRIA ELETRÔNICA UTILIZANDO RECONHECIMENTO DE FALA"

Fundamentos

5 A discussão abaixo é meramente fornecida para informação de suporte geral e não deve ser utilizada como um auxílio na determinação do escopo da matéria reivindicada.

O uso de sistemas de chamada telefônica de saída automatizados é bem comum. Além de serem utilizados para fins de comercialização, tais sistemas também podem ser utilizados por consultórios médicos ou clínicas para contatar pacientes para confirmar ou alterar suas consultas, por escolas para informar aos alunos sobre mudanças de programação, por entidades de caridade para obtenção de contribuições, e agências governamentais para fornecer a notificação ou outras informações, para citar apenas algumas poucas das outras aplicações.

Em muitos casos, pode ser necessário ou útil se determinar automaticamente se o recipiente da chamada telefônica é uma pessoa de verdade ou uma secretária eletrônica. Dependendo de se uma pessoa atendeu ou uma secretária eletrônica está em uso, diferentes ações podem ser tomadas pelo sistema de chamada telefônica de saída. No entanto, essa tarefa, análise de chamada, é difícil e atualmente imprecisa.

A análise de chamada é comumente realizada no nível de comutação de hardware. A análise é implementada pela utilização de um intervalo curto quando o recipiente inicialmente atende à chamada e antes de a chamada ser conectada ao aplicativo de telefonia. Durante esse intervalo, quando o recipiente começa a falar, o sistema processará os sinais audíveis recebidos com relação a, por exemplo, conteúdo de energia, intensidade ou outros parâmetros de sinal de sinais audíveis, a fim de determinar se o recipiente é uma pessoa ou uma secretária eletrônica. É importante se compreender que o aplicativo de telefonia não sabe que a chamada foi atendida nesse momento, e, portanto, não distribuiu qualquer aviso inicial. Dessa forma, na outra extremidade da linha, apesar de o recipiente ter atendido à chamada e ter dito uma saudação tipo "alô", o recipiente só ouve silêncio de volta visto que o sistema está realizando a análise da chamada. Em muitos casos, o recipiente simplesmente desligará.

30 Sumário

O Sumário é fornecido para introduzir uma seleção de conceitos de uma forma simplificada que são adicionalmente descritos abaixo na Descrição Detalhada. O Sumário não deve identificar características chave ou características essenciais da matéria reivindicada, nem deve ser utilizado como um auxílio na determinação do escopo da matéria reivindicada. Adicionalmente, a matéria reivindicada não está limitada às implementações que solucionam qualquer ou todas as desvantagens notadas nos fundamentos.

Um módulo de detecção de secretária eletrônica é utilizado para determinar se um

recipiente de chamada é uma pessoal ou uma secretária eletrônica. O módulo de detecção de secretária eletrônica inclui um reconhecedor de fala e um módulo de análise de chamada. O reconhecedor de fala recebe uma resposta audível do recipiente de chamada para uma chamada. O reconhecedor de fala processa a resposta audível e fornece uma saída indicativa da fala reconhecida. O módulo de análise de chamada processa a saída do reconhecedor de fala para gerar uma saída indicativa de se o recipiente da chamada é uma pessoa ou uma secretária eletrônica. Em uma modalidade, o módulo de análise de chamada pode incluir um módulo classificador que fornece análise estatística da saída do reconhecedor de fala para determinar se um recipiente de chamada é uma pessoa ou secretária eletrônica.

Também é descrita uma técnica para garantir que toda a mensagem do chamador seja gravada pela secretária eletrônica. Em particular, um reconhecedor de fala é operado para detectar eventos de "invasão" pela secretária eletrônica, e onde a mensagem é reproduzida para a secretária eletrônica se um evento de "invasão" for detectado. Apesar disso poder fazer com que a mensagem seja reproduzida uma ou mais vezes, isso é particularmente vantajoso visto que qualquer evento de "invasão" significa que a saudação da secretária eletrônica ainda não acabou, e, dessa forma, a secretária eletrônica não está pronta para gravar uma mensagem. Pela reprodução da mensagem após cada evento de "invasão", depois da reprodução da mensagem após o último evento de "invasão" é garantido que toda a mensagem seja gravada quando a secretária eletrônica está pronta para gravar a mensagem.

Breve Descrição dos Desenhos

A figura 1 é uma representação em diagrama em bloco esquemática de um sistema de discagem;

A figura 2 é uma representação em diagrama em bloco esquemática de componentes do sistema de discagem ilustrado na figura 1;

A figura 3 é um diagrama em bloco esquemático para o processamento de uma chamada;

A figura 4 é um fluxograma para o processamento de uma chamada;

A figura 5 é um fluxograma mais detalhado para o processamento da chamada;

A figura 6 é um ambiente de computação ilustrativo.

Descrição Detalhada

As figuras 1 e 2 são diagramas em bloco esquemáticos de um sistema de discagem ou de chamada de saída 100. O sistema 100 é implementado de forma ilustrativa dentro de um dos ambientes de computação discutidos em associação com a figura 6 discutida abaixo, ou dentro de outros ambientes de computação adequados. Como ilustrado na figura 1, o sistema de chamada de saída 100 realiza as chamadas para qualquer um dentre a plurali-

dade de recipientes de chamada 102 (recipientes de chamada 102-1 a 102-M são ilustrados). Como utilizado aqui, um "recipiente de chamada" pode ser uma pessoa ou uma secretária eletrônica. Depois da realização de uma chamada para um dos recipientes de chamada 102-1 a 102-M, o sistema de chamada de saída 100 determina se o recipiente da chamada é uma pessoa ou se uma secretária eletrônica está em uso.

A figura 2 ilustra em maiores detalhes os componentes ou módulos do sistema de chamada de saída 100 de acordo com uma modalidade ilustrativa. Como ilustrado, o sistema de chamada de saída 100 realiza as chamadas para um ou mais recipientes de chamada 102. Um módulo de detecção de secretária eletrônica 106 é utilizado para determinar se o recipiente da chamada 102 é uma pessoa ou uma secretária eletrônica. Em uma aplicação típica, um módulo de aplicativo 104 inicia a discagem do número de telefone do recipiente da chamada recuperado a partir de, por exemplo, uma base de dados de números telefônicos 108. O módulo de detecção de secretária eletrônica 106 é então utilizado para determinar se o recipiente da chamada 102 é uma pessoa ou uma secretária eletrônica. Ações diferentes serão tomadas pelo sistema de chamada de saída 100 de acordo com a saída do módulo de detecção 106. Por exemplo, se o recipiente da chamada for uma pessoa, o módulo de aplicativo 104 pode então transferir a chamada para um operador 110, ou engatar o recipiente em uma conversa automática. Do contrário, o módulo de aplicação 104 pode ser utilizado para deixar uma mensagem na secretária eletrônica do recipiente da chamada 102. Na figura 2, os recipientes da chamada 102, a base de dados do número telefônico 108 e o operador 110 são ilustrados por linhas tracejadas para ilustrar que os mesmos não fazem parte de, ou são necessários no sistema de chamada de saída 100. Adicionalmente, o diagrama em bloco do sistema de chamada de saída 100 é esquemático para fins de compreensão e a fim de fornecer uma relação ilustrativa entre o módulo de detecção de secretária eletrônica 106 e seus componentes, além de outros componentes do sistema de chamada de saída 100. Deve-se compreender que essa é apenas uma modalidade ilustrativa visto que um ou mais componentes do sistema de chamada de saída 100 e módulo de detecção de secretária eletrônica 106 podem ser omitidos e/ou combinados em outras configurações em outras modalidades.

Na modalidade ilustrada, o módulo de detecção de secretária eletrônica 106 inclui um reconhecedor de fala 120 e um módulo de análise de chamada 122. As figuras 3 e 4 são um diagrama esquemático e um fluxograma de um método 200, respectivamente, ilustrando a operação desses módulos. Na etapa 202, o reconhecedor de fala 120 recebe a resposta audível 124 do recipiente de chamada 102 para a chamada realizada pelo sistema de chamada de saída 100. O reconhecedor de fala 120 processa a resposta audível 124 e fornece uma saída 126 indicativa da fala reconhecida, e em uma modalidade, frases, comumente na forma de dados de texto. Como utilizado aqui uma frase é uma pluralidade de palavras no

contexto como fornecido pelo recipiente da chamada.

Na etapa 204, a saída 126 do reconhecedor de fala 120 é fornecida como entrada para o módulo de análise de chamada 122. O módulo de análise de chamada 122 processa a saída 126 do reconhecedor de fala 120 para gerar uma saída 128 indicativa de se o recipiente da chamada é uma pessoa ou uma secretária eletrônica. O módulo de análise de chamada 122 pode incluir um módulo classificador 123 que fornece análise estatística do conteúdo das frases na saída 126. O reconhecedor de fala 120 e o módulo classificador 123 são módulos bem conhecidos e podem ser implementados utilizando-se uma variedade de técnicas bem conhecidas. No entanto, deve-se notar que um modelo de linguagem 130 (por exemplo, N-gram, gramática livre de contexto, híbrida, etc.) utilizado pelo reconhecedor de fala 120 e um modelo classificador 132 são tipicamente ambos seqüenciados em apenas frases ou saudações utilizadas por pessoas e secretárias eletrônicas quando atendem uma chamada. Por exemplo, uma pessoa pode atender a uma chamada telefônica com a saudação "Alô", "Como posso ajudar?", "Aqui quem fala é o Steve" ou talvez dizendo apenas o nome. Em contraste, uma secretária eletrônica pode ter a saudação "Você ligou para a secretária eletrônica. Não estou disponível no momento. Por favor, deixe sua mensagem." ou simplesmente "Favor deixar uma mensagem". No entanto, na detecção de secretária eletrônica com base em conteúdo, pode ser necessário se reconhecer frases importantes tal como "não disponível no momento" ou "deixe uma mensagem". Da mesma forma, no caso de passar por um sistema de identificação de telefone no local de recepção, pode ser necessário se reconhecer frases como "pressione 2". Para uma precisão de reconhecimento alta das frases como essa, o modelo de linguagem 130 pode ser seqüenciado com frases importantes e onde o modelo pode ser suavizado com um modelo de enchimento n-gram pra capturar as palavras não cobertas nas frases importantes. Um exemplo de um modelo de enchimento n-gram é descrito por D.Yu, Y. - C, Ju, Y. Wang e A. Acero em "N-gram Based Filler Model for Robust Grammar Authoring," publicado em Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, maio de 2006. Utilizando um número suficiente de exemplos (exemplos genéricos ou exemplos específicos de aplicativo) ambos o modelo de linguagem 130 e o modelo classificador 132 podem ser seqüenciados de forma adequada.

Com relação ao módulo de análise de chamada 122 e etapa 204, as características de não fala 136 também podem ser utilizadas para análise em adição ou na alternativa à saída 126 do reconhecedor de fala 120. Exemplos de características de não fala 136 incluem, mas não estão limitados a: se o recipiente de chamada 102 "invadiu" (isso é, interrompeu um aviso sendo exibido quando o aplicativo 104 está executando um diálogo), a duração da resposta audível feita pelo recipiente da chamada 102 quando atende a chamada, e se ou não o reconhecedor de fala 120 foi capaz de reconhecer a resposta audível 124 como

uma frase válida. Na figura 3, as características de não fala 136 são produzidas a partir de um gerador de característica de não palavra 128, que pode receber e processar a resposta audível 124 ou outra informação de entrada 138 pertencente à chamada do sistema de chamada de saída 100. Na figura 4, o recebimento das características de não palavra 136 é indicado na etapa 203.

Nesse ponto, deve-se notar também que não existe qualquer exigência de que o reconhecedor de fala 120 seja capaz de reconhecer a resposta audível completa 124, mas ao invés disso, devido ao ambiente barulhento no recipiente da chamada 102, ao ruído em uma saudação gravada ou ao ruído do sistema de telefonia, e à natureza da resposta, apenas uma ou mais partes da resposta audível 124 podem ser reconhecíveis, e, dessa forma, utilizadas para determinar se o recipiente da chamada 102 é uma pessoa ou uma secretária eletrônica.

A figura 5 é um fluxograma ilustrando um método ilustrativo 300 realizado pelo módulo de análise de chamada 122 na etapa de execução 204 na figura 3. A etapa 302 representa o começo do reconhecimento da fala depois de a chamada ter sido realizada. Uma primeira característica de não fala que pode opcionalmente ser utilizada para a determinação de se o recipiente é uma pessoa ou uma secretária eletrônica é se o silêncio está inicialmente presente uma vez que a chamada é atendida. Em particular, descobriu-se que uma pessoa pode esperar, por exemplo, por um ou dois segundos, antes de começar a falar. Dessa forma, se na etapa 304 for determinado que o silêncio significativo está presente depois de a chamada ter sido atendida, por exemplo, como medido por um temporizador excedendo um tempo selecionado a partir da etapa 302, uma determinação pode ser feita de que o recipiente da chamada 102 é uma pessoa.

Se, por outro lado, o silêncio significativo não estiver presente depois de a chamada ter sido atendida na etapa 302, ou a etapa 302 não estiver presente, a duração da resposta audível pode ser utilizada para determinar se ou não o recipiente da chamada é uma pessoa. As etapas 306 e 308 ilustram o processamento da característica de não palavra compreendendo a duração da resposta audível. Na etapa 306, se a duração da resposta audível for muito curta (por exemplo, de aproximadamente menos de 1 segundo), é mais do que provável que uma pessoa tenha atendido à chamada. Em contraste, se a duração da resposta audível for muito longa (por exemplo, de aproximadamente quatro ou mais segundos) como ilustrado na etapa 308, o recipiente da chamada é provavelmente uma secretária eletrônica. Deve-se notar que a ordem na qual as etapas 306 e 308 são ilustradas não é necessária visto que essas etapas podem ser invertidas. Da mesma forma, os períodos de tempo especificados podem ser ajustados.

Se uma determinação de se o recipiente da chamada é uma pessoa ou uma secretária eletrônica não tiver sido realizada antes da etapa 308, o fluxo de processo continua

para a etapa 310 onde a duração da resposta audível pode novamente ser utilizada. Em particular, na etapa 310 a duração da resposta audível é medida para determinar se é relativamente curta, por exemplo, dois ou três segundos. O resultado dessa etapa é combinado com a informação relacionada com o contexto da resposta audível como reconhecida pelo reconhecedor de fala 120 tal como obtido a partir do módulo classificador 123. Em outras palavras, o módulo classificador 123 analisa a saída 126 do reconhecedor de fala 120 para determinar se uma ou mais frases estão estatisticamente consistentes com as frases utilizadas por uma pessoa ou frases utilizadas por uma secretária eletrônica. Na figura 5, a informação do módulo classificador 123 é representada nas etapas 312 e 314. Em particular, se na etapa 312 a informação do módulo classificador 123 indicar que as frases na resposta audível são muito provavelmente (alta confiança) de uma pessoa, a saída 128 do módulo de análise de chamada 122 indicará que o recipiente da chamada 102 é uma pessoa. Em contraste, se na etapa 312 não houver alta confiança de o recipiente da chamada ser uma pessoa, a saída do módulo de análise de chamada 122 indicará que o recipiente da chamada 102 é uma secretária eletrônica. De forma similar, na etapa 314, se a informação do módulo classificador 123 indicar que as frases na resposta audível são muito provavelmente de uma secretária eletrônica, a saída do módulo de análise de chamada 122 indicará que o recipiente da chamada é uma secretária eletrônica, ao passo que se alta confiança de que o recipiente da chamada é uma secretária eletrônica não estiver presente, a saída do módulo de análise de chamada 122 indicará que o recipiente da chamada 102 é uma pessoa. Deve-se notar que nesse exemplo existe uma tendência ao erro no lado da classificação errada de um recipiente de chamada como uma pessoa ao invés de uma secretária eletrônica. Se for desejável, a tendência pode ser no sentido da classificação do recipiente de chamada como uma secretária eletrônica, ou nenhuma tendência de acordo com os custos associados com diferentes erros.

Favor notar que o classificador estatístico descrito acima é apenas um exemplo ilustrativo. O módulo de análise de chamada 122 pode utilizar muitos outros classificadores, tal como classificadores de entropia máxima. Adicionalmente, o módulo de análise de chamada 122 também pode utilizar máquinas de vetor de suporte, componentes utilizando árvores de decisão, e redes neurais artificiais para alcançar uma precisão comparável.

No exemplo da figura 5, as características de não palavra tal como a presença de silêncio depois de a chamada ter sido atendida pelo recipiente e a duração da resposta audível foram utilizadas antes da classificação real do conteúdo da resposta audível a fim de determinar se o recipiente de chamada é uma pessoa ou secretária eletrônica. Novamente, outras características de não palavra tal como se o recipiente invadiu ou se o reconhecedor de fala 120 foi capaz de reconhecer a resposta audível também podem ser utilizadas. Adicionalmente, um classificador pode ser seqüenciado com base nas características de não

palavra e utilizado com o classificador 123 fornecendo uma saída com base nas frases reconhecidas da resposta audível para fornecer uma indicação de se o recipiente da chamada é uma pessoa ou uma secretária eletrônica. Da mesma forma, um classificador único seria seqüenciado com base nas características de não palavra e as palavras reconhecidas da resposta audível para fornecer uma indicação de se o recipiente da chamada é uma pessoal ou uma secretária eletrônica.

Em algumas aplicações, pode ser necessário se reproduzir uma mensagem para o recipiente da chamada, por exemplo, se tiver sido determinado que o recipiente da chamada é uma secretária eletrônica. Apesar de as secretárias eletrônicas fornecerem um tom ou silêncio significando que a saudação terminou e que uma mensagem deve agora ser deixada, o reconhecimento desse tom é difícil. Em muitos casos, visto que o tom ou silêncio pode não ser reconhecido com precisão, pelo menos parte da mensagem pode ser reproduzida durante a saudação, dessa forma, a parte inicial da mensagem pode não ser gravada.

Uma etapa opcional 206 ilustrada na figura 4, compreendendo também outro aspecto descrito aqui fornece meios e uma técnica para se garantir que toda a mensagem seja gravada pela secretária eletrônica ou, se for desejável, ouvida pela pessoa. Em particular, a mensagem é reproduzida com esse reconhecedor de fala 120 ligado e ativado para "invasão" (de forma a detectar a finalização da saudação) até que a mensagem tenha sido completada. Por exemplo, assumindo-se que uma secretária eletrônica forneça uma saudação "Você ligou para....Eric Moe....Favor deixar uma mensagem," onde "..." representa silêncio durante a saudação. Em muitos sistemas atuais, a mensagem pode ser reproduzida depois da detecção de um desses períodos de silêncio, onde a mensagem pode então ser reproduzida, mas não gravada visto que a secretária eletrônica não alcançou o final da saudação. Utilizando-se a técnica descrita aqui, o reconhecedor de fala 120 com capacidade de "invasão" detecta cada parte da saudação e faz com que a mensagem seja iniciada novamente. Uma vez que a saudação alcançou definitivamente seu final, a mensagem é reproduzida, mas agora até o seu final de modo a garantir que a mensagem seja ouvida ou gravada em sua totalidade.

A figura 6 ilustra um exemplo de um ambiente de sistema de computação adequado 400 no qual os conceitos descritos aqui podem ser implementados. Em particular, uma forma do ambiente de sistema de computação 400 pode ser utilizada para o sistema de chamada de saída 100 ou partes do mesmo. Não obstante, o ambiente de sistema de computação 400 é novamente apenas um exemplo de um ambiente de computação adequado para cada um desses computadores e não deve sugerir qualquer limitação quanto ao escopo de uso ou funcionalidade da descrição abaixo. Nem deve o ambiente de computação 400 ser interpretado como sendo dependente ou possuindo qualquer exigência com relação a qualquer um ou uma combinação de componentes ilustrados no ambiente operacional ilustrativo

400.

Em adição aos exemplos fornecidos aqui, outros sistemas de computação bem conhecidos, ambientes, e/ou configurações podem ser adequados para uso com os conceitos descritos aqui. Tais sistemas incluem, mas não estão limitados a, computadores pessoais, computadores servidores, dispositivos portáteis, sistemas multiprocessadores, sistemas com Bse em micro processador, aparelhos decodificadores, partes eletrônicas de consumidor programáveis, PCs em rede, mini computadores, computadores principais, ambientes de computação distribuídos que incluem qualquer um dos sistemas acima ou dispositivos, e similares.

Os conceitos descritos aqui podem ser consubstanciados no contexto geral das instruções executáveis por computador, tal como módulos de programa, sendo executados por um computador. Geralmente, os módulos de programa incluem rotinas, programas, objetos, componentes, estruturas de dados, etc. que realizam tarefas particulares ou implementam tipos de dados abstratos particulares. Os versados na técnica podem implementar a descrição e/ou figuras aqui como instruções executáveis por computador, que podem ser consubstanciadas em qualquer forma de mídia legível por computador discutida abaixo.

Os conceitos descritos aqui também podem ser praticados em ambientes de computação distribuídos onde as tarefas são realizadas por dispositivos de processamento remoto que são conectados através de uma rede de comunicações. Em um ambiente de computação distribuída, os módulos de programa podem ser localizados em ambas a mídia de armazenamento em computador local e remota incluindo dispositivos de armazenamento em memória.

Com referência à figura 6, um sistema ilustrativo inclui um dispositivo de computação de finalidade geral na forma de um computador 410. Os componentes de computador 410 podem incluir, mas não estão limitados a, uma unidade de processamento 420, uma memória de sistema 430 e um barramento de sistema 421 que acopla vários componentes de sistema incluindo a memória do sistema à unidade de processamento 420. O barramento do sistema 421 pode ser qualquer um dentre os vários tipos de estruturas de barramento incluindo o barramento de memória ou controlador de memória, um barramento periférico, e um barramento local utilizando qualquer uma dentre uma variedade de arquiteturas de barramento. Por meio de exemplo, e não de limitação, tais arquiteturas incluem o barramento de Arquitetura Padrão (ISA), o barramento de Arquitetura de Micro Canal (MCA), o barramento ISA Melhorado (EISA), o barramento local da Associação de Padrões Eletrônicos de Vídeo (VESA), o barramento de Interconexão de Componente Periférico (PCI) também conhecido como barramento Mezzanine.

O computador 410 inclui tipicamente uma variedade de mídias legíveis por computador. A mídia legível por computador pode ser qualquer mídia disponível que possa ser

acessada pelo computador 410 e inclui ambas as mídias volátil e não volátil, e as mídias removível e não removível. Por meio de exemplo, e não de limitação, a mídia legível por computador pode compreender mídia de armazenamento em computador. Por meio de exemplo, e não de limitação, a mídia legível por computador pode compreender mídia de armazenamento em computador. A mídia de armazenamento em computador inclui mídia volátil e não volátil, removível e não removível implementada em qualquer método ou tecnologia para armazenamento de informação tal como instruções legíveis por computador, estruturas de dados, módulos de programa ou outros dados. A mídia de armazenamento em computador inclui, mas não está limitada a, memórias RAM, ROM, EEPROM e FLASH ou outra tecnologia de memória, CD-ROM, discos versáteis digitais (DVD) ou outro armazenamento em disco ótico, fitas magnéticas, armazenamento em disco magnético ou outros dispositivos de armazenamento magnético, ou qualquer outro meio que possa ser utilizado para armazenar a informação desejada e que possa ser acessado pelo computador 400.

A memória de sistema 430 inclui a mídia de armazenamento em computador na forma de memória volátil e/ou não volátil tal como memória de leitura apenas (ROM) 431 e memória de acesso randômico (RAM) 432. Um sistema básico de entrada/saída (BIOS) 433, contendo as rotinas básicas que ajudam a transferir informação entre os elementos dentro do computador 410, tal como durante a inicialização, é tipicamente armazenado em ROM 431. A RAM 432 contém tipicamente dados e/ou módulos de programa imediatamente acessíveis a e/ou atualmente sendo operados pela unidade de processamento 420. Por meio de exemplo, e não de limitação,

A figura 6 ilustra o sistema operacional 434, programas de aplicativo 435, outros módulos de programa 436, e dados de programa 437. Aqui, os programas de aplicativo 435, os módulos de programa 436 e os dados de programa 437 implementam um ou mais dos conceitos descritos acima.

O computador 410 também pode incluir outras mídias de armazenamento em computador removíveis/não removíveis, voláteis/não voláteis. Por meio de exemplo apenas, a figura 6 ilustra um acionador de disco rígido 441 que lê a partir de ou escreve na mídia magnética não removível e não volátil, um acionador de disco magnético 451 que lê a partir de ou escreve em um disco magnético não volátil e removível 452, e um acionador de disco ótico 455 que lê a partir de ou escreve em um disco ótico removível e não volátil 456 tal como um CD ROM ou outra mídia ótica. Outras mídias de armazenamento em computador removíveis/não removíveis e voláteis/não voláteis que podem ser utilizadas no ambiente operacional ilustrativo incluem, mas não estão limitadas a, fitas magnéticas, cartões de memória flash, discos versáteis digitais, fita de vídeo digital, RAM em estado sólido, ROM em estado sólido, e similares. O acionador de disco rígido 441 é tipicamente conectado ao barramento do sistema 421 através de uma interface de memória não removível tal como a in-

terface 440, e o acionador de disco magnético 451 e o acionador de disco ótico 455 são tipicamente conectados ao barramento do sistema 421 por uma interface de memória removível, tal como a interface 450.

Os acionadores e sua mídia de armazenamento em computador associada discutidos acima e ilustrados na figura 6 fornecem armazenamento de instruções legíveis por computador, estruturas de dados, módulos de programa e outros dados para o computador 410. Na figura 6, por exemplo, o acionador de disco rígido 441 é ilustrado como armazenando o sistema operacional 444, os programas de aplicativo 445, outros módulos de programa 446, e dados de programa 447. Note-se que esses componentes podem ser iguais ou diferentes do sistema operacional 434, programas de aplicativo 435, outros módulos de programa 436 e dados de programa 437. O sistema operacional 444, os programas de aplicativo 445, outros módulos de programa 446 e os dados de programa 447 recebem números diferentes para ilustrar que, no mínimo, são cópias diferentes.

Um usuário pode registrar os comandos e informação no computador 410 através dos dispositivos de registro tal como um teclado 462, um microfone 463 e um dispositivo de apontar 461, tal como um mouse, TrackBall ou teclado de toque. Esses e outros dispositivos de registro são freqüentemente conectados à unidade de processamento 420 através de uma interface de registro de usuário 460 que é acoplada ao barramento do sistema, mas podem ser conectados por outra interface e estruturas de barramento, tal como uma porta serial ou um barramento serial universal (USB). Um monitor 491 ou outro tipo de dispositivo de exibição também é conectado ao barramento do sistema 421 através de uma interface, tal como uma interface de vídeo 490.

O computador 410 pode operar em um ambiente em rede utilizando conexões lógicas com um ou mais computadores remotos, tal como um computador remoto 480. O computador remoto 480 pode ser um computador pessoal, um dispositivo portátil, um servidor, um roteador, um PC em rede, um dispositivo não hierarquizado ou outro nó de rede comum, e inclui tipicamente muitos ou todos os elementos descritos acima com relação ao computador 410. As conexões lógicas apresentadas na figura 6 incluem uma rede de área local (LAN) 471 e uma rede de área ampla (WAN) 473, mas também podem incluir outras redes. Tais ambientes em rede são lugar comum em escritórios, redes de computador empresariais, intranets e na Internet.

Quando utilizado em um ambiente em rede LAN, o computador 410 é conectado à LAN 471 através de uma interface de rede ou adaptador 470. Quando utilizado em um ambiente em rede WAN, o computador 410 tipicamente inclui um modem 472 ou outro meio de estabelecimento de comunicações através da WAN 473, tal como a Internet. O modem 472, que pode ser interno ou externo, pode ser conectado ao barramento do sistema 421 através da interface de registro de usuário 460, ou outro mecanismo adequado. Em um ambiente de

rede, os módulos de programa apresentados com relação ao computador 410, ou partes dos mesmos, podem ser armazenados no dispositivo de armazenamento em memória remota. Por meio de exemplo, e não de limitação, a figura 6 ilustra programas de aplicativo remotos 485 residindo no computador remoto 480. Será apreciado que as conexões de rede ilustradas são ilustrativas e que outros meios de estabelecimento de uma conexão de comunicação entre os computadores podem ser utilizados.

Deve-se notar que os conceitos descritos aqui podem ser realizados em um sistema de computador tal como o descrito com relação à figura 6. No entanto, outros sistemas adequados incluem um servidor, um computador devotado ao manuseio de mensagens, ou em um sistema distribuído no qual partes diferentes dos conceitos são realizadas em diferentes partes do sistema de computação distribuído.

Apesar de a matéria ter sido descrito em linguagem específica de características estruturais e/ou atos metodológicos, deve-se compreender que a matéria definida nas reivindicações em anexo não está limitada às características específicas ou atos descritos acima como foi determinado nos tribunais. Ao invés disso, as características específicas e atos descritos acima são descritos como formas ilustrativas de implementação das reivindicações.

REIVINDICAÇÕES

1. Método implementado por computador (200) de determinação de se um recipiente de chamada é uma pessoa ou uma secretária eletrônica, o método sendo **CARACTERIZADO** pelo fato de compreender:

5 o recebimento (202) de uma resposta audível de um recipiente de chamada e o processamento da resposta audível com um reconhecedor de fala para fornecer uma saída indicativa da fala reconhecida; e

 o processamento (204) da saída indicativa de fala reconhecida com um classificador para fornecer uma saída indicativa de se o recipiente da chamada é uma pessoa ou uma
10 secretária eletrônica, o dito processamento sendo baseado em análise estatística da saída indicativa de fala reconhecida.

2. Método implementado por computador (200), de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de o processamento (204) incluir o processamento da resposta audível para determinar as características de não palavra associadas com a resposta audível (304, 306, 308) e utilizar as características de não palavra como base para fornecer uma
15 saída indicativa de se o recipiente da chamada é uma pessoa ou uma secretária eletrônica.

3. Método implementado por computador (200), de acordo com a reivindicação 2, **CARACTERIZADO** pelo fato de o processamento (204) incluir a determinação de uma duração de resposta audível (306, 308) e utilização da duração da resposta audível como uma
20 base para fornecer uma saída indicativa de se o recipiente da chamada é uma pessoa ou uma secretária eletrônica.

4. Método implementado por computador (200), de acordo com a reivindicação 2, **CARACTERIZADO** pelo fato de o processamento (204) incluir a determinação de se um silêncio está inicialmente presente na resposta audível (304) e a utilização da presença do
25 silêncio como base para fornecer uma saída indicativa de se o recipiente da chamada é uma pessoa ou uma secretária eletrônica.

5. Método implementado por computador (200), de acordo com a reivindicação 4, **CARACTERIZADO** pelo fato de o processamento (204) incluir a determinação de se o silêncio por um período relativo a um comprimento selecionado está inicialmente presente na
30 resposta audível (304) e a utilização da presença do silêncio como base para fornecer uma saída indicativa de se o recipiente da chamada é uma pessoal ou uma secretária eletrônica.

6. Método implementado por computador (200), de acordo com a reivindicação 2, **CARACTERIZADO** pelo fato de o processamento (204) incluir a utilização de uma indicação de se o reconhecedor de fala foi capaz de reconhecer qualquer fala na resposta audível como uma base para fornecer uma saída indicativa de se o recipiente da chamada é uma pessoal ou uma secretária eletrônica.
35

7. Método implementado por computador (200), de acordo com a reivindicação 3,

CARACTERIZADO pelo fato de o processamento (204) incluir a utilização de análise estatística (312, 314) quando a duração da fala está dentro de uma faixa selecionada.

8. Método implementado por computador (200), de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de o processamento (204) incluir a utilização da análise estatística (312, 314) quando uma duração de fala está dentro de uma faixa selecionada.

9. Meio legível por computador possuindo instruções que quando implementadas por um computador determinam se um recipiente de chamada é uma pessoal ou uma secretária eletrônica, as instruções sendo **CARACTERIZADAS** pelo fato de compreender:

o recebimento (202) de uma resposta audível de um recipiente de chamada e o processamento da resposta audível com um reconhecedor de fala para fornecer uma saída indicativa da fala reconhecida; e

o processamento (204) da saída indicativa de fala reconhecida para fornecer uma saída indicativa de se o recipiente da chamada é uma pessoal ou uma secretária eletrônica, o dito processamento sendo baseado em análise de frases de palavra na saída indicativa da fala reconhecida.

10. Meio legível por computador, de acordo com a reivindicação 9, **CARACTERIZADO** pelo fato de o processamento (204) compreender o processamento (312, 314) das frases de palavra com um classificador estatístico e a utilização de uma saída do classificador estatístico como uma base para fornecer uma saída indicativa de se o recipiente da chamada é uma pessoal ou uma secretária eletrônica.

11. Meio legível por computador, de acordo com a reivindicação 9, **CARACTERIZADO** pelo fato de o processamento (204) incluir o processamento da resposta audível para determinar as características de não palavra associadas com a resposta audível (304, 306, 308) e a utilização das não características como uma base para fornecer uma saída indicativa de se o recipiente da chamada é uma pessoal ou uma secretária eletrônica.

12. Meio legível por computador, de acordo com a reivindicação 9, **CARACTERIZADO** pelo fato de o processamento incluir a determinação de uma duração de resposta audível (306, 308) e a utilização da duração da resposta audível como uma base para fornecer uma saída indicativa de se o recipiente da chamada é uma pessoal ou uma secretária eletrônica.

13. Meio legível por computador, de acordo com a reivindicação 9, **CARACTERIZADO** pelo fato de o processamento incluir a determinação de se o silêncio está inicialmente presente na resposta audível (304) e a utilização da presença do silêncio como uma base para fornecer uma saída indicativa de se o recipiente da chamada é uma pessoal real ou uma secretária eletrônica.

14. Meio legível por computador, de acordo com a reivindicação 13, **CARACTERIZADO** pelo fato de o processamento (204) incluir a determinação de se o si-

lêncio de um período relativo a um comprimento selecionado está inicialmente presente na resposta audível (304) e a utilização da presença do silêncio como uma base para fornecer uma saída indicativa de se o recipiente da chamada é uma pessoal ou uma secretária eletrônica.

5 15. Meio legível por computador, de acordo com a reivindicação 9, **CARACTERIZADO** pelo fato de o processamento (204) incluir a utilização de uma indicação de se o reconhecedor de fala foi capaz de reconhecer qualquer fala na resposta audível como uma base para fornecer uma saída indicativa de se o recipiente da chamada é uma pessoal ou uma secretária eletrônica.

10 16. Meio legível por computador, de acordo com a reivindicação 10, **CARACTERIZADO** pelo fato de o processamento (204) incluir a utilização de análise estatística quando a duração da fala está dentro de uma faixa selecionada.

15 17. Meio legível por computador, de acordo com a reivindicação 10, **CARACTERIZADO** pelo fato de compreender adicionalmente a operação (206) do reconhecedor de fala para detectar eventos de invasão pelo recipiente da chamada e reproduzir uma mensagem para o recipiente da chamada se um evento de invasão for detectado até que a mensagem seja reproduzida em sua totalidade.

20 18. Meio legível por computador, de acordo com a reivindicação 17, **CARACTERIZADO** pelo fato de a reprodução da mensagem para o recipiente da chamada (206) compreender a reprodução da mensagem para o recipiente da chamada se o recipiente da chamada for uma secretária eletrônica.

19. Método implementado por computador para deixar uma mensagem em uma secretária eletrônica, o método sendo **CARACTERIZADO** pelo fato de compreender:

25 a operação (206) de um reconhecedor de fala para detectar eventos de invasão pela secretária eletrônica; e

a reiniciação (206) de uma mensagem para a secretária eletrônica se um evento de invasão for detectado até que a mensagem seja reproduzida em sua totalidade.

30 20. Método implementado por computador, de acordo com a reivindicação 19, **CARACTERIZADO** pelo fato de compreender adicionalmente, antes da etapa de reprodução de mensagem, a detecção de se um recipiente de chamada é uma pessoal ou uma secretária eletrônica onde a detecção compreende:

o recebimento (202) de uma resposta audível do recipiente da chamada e o processamento da resposta audível com o reconhecedor de fala para fornecer uma saída indicativa da fala reconhecida; e

35 o processamento (204) da saída indicativa da fala reconhecida para fornecer uma saída indicativa de se o recipiente da chamada é uma pessoa ou uma secretária eletrônica, o dito processamento sendo baseado na análise das frases de palavras na saída indicativa

de fala reconhecida.

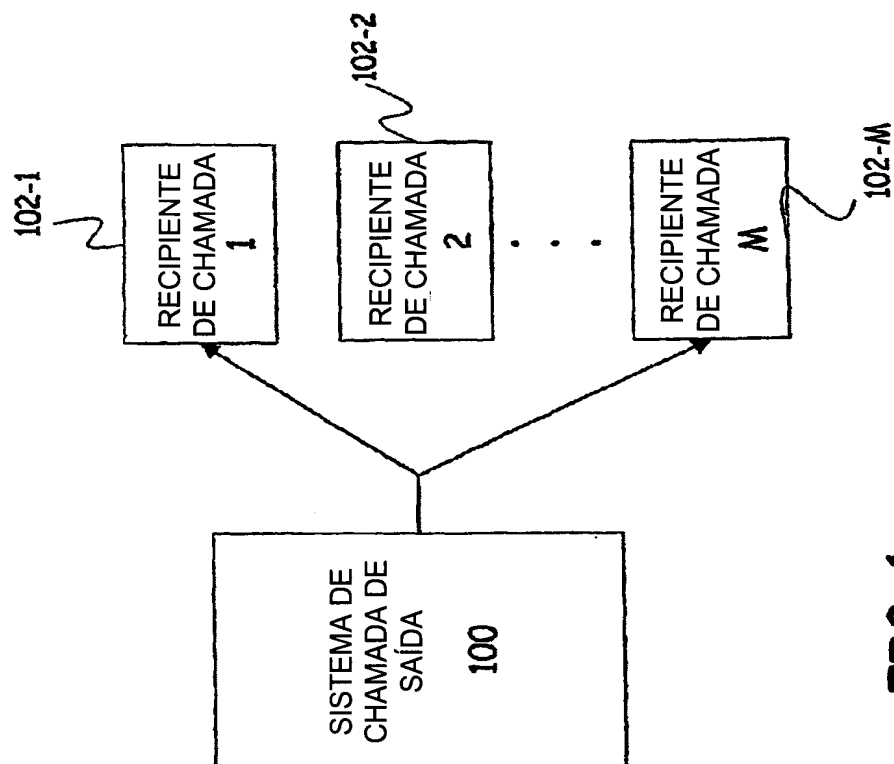
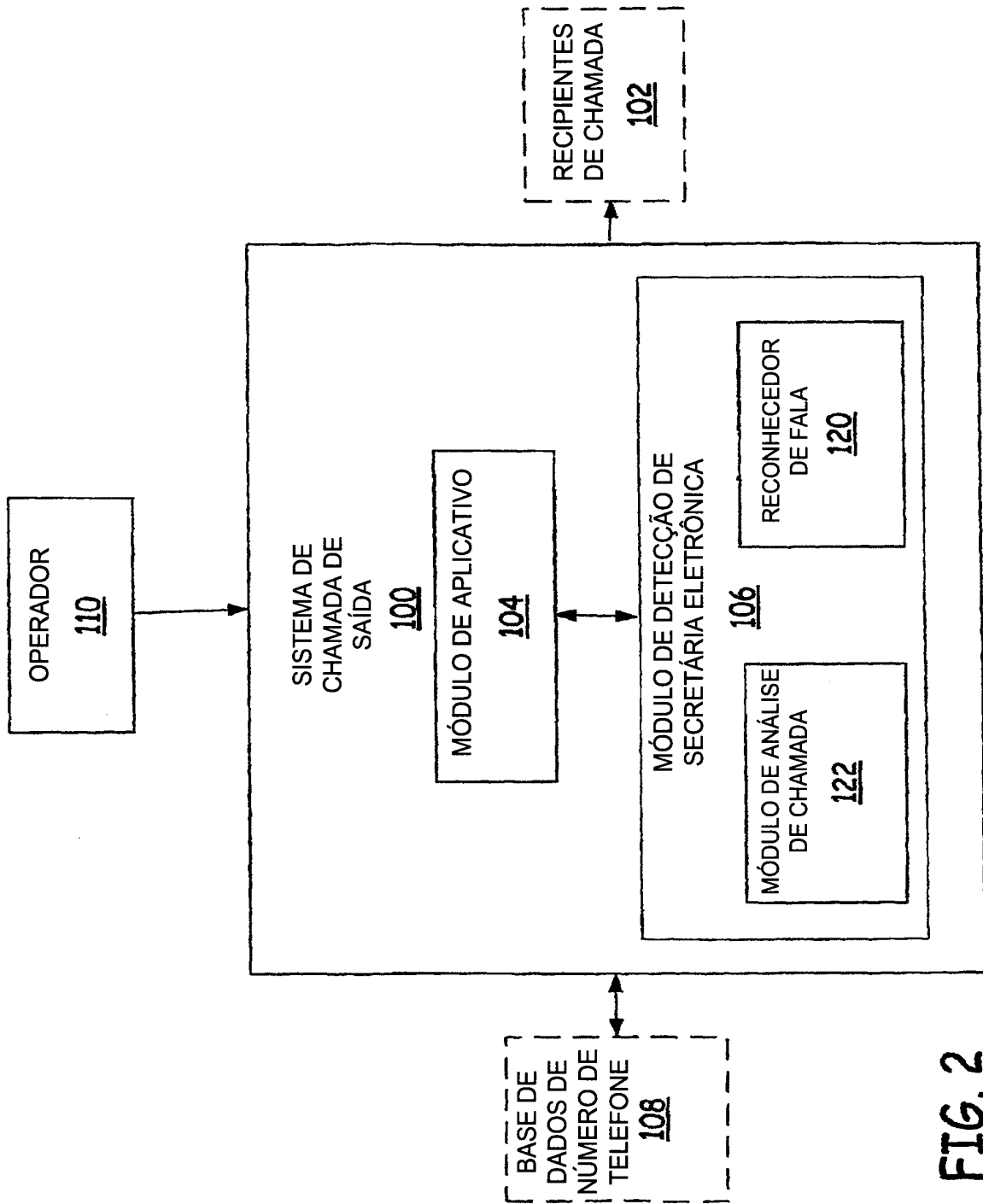


FIG. 1



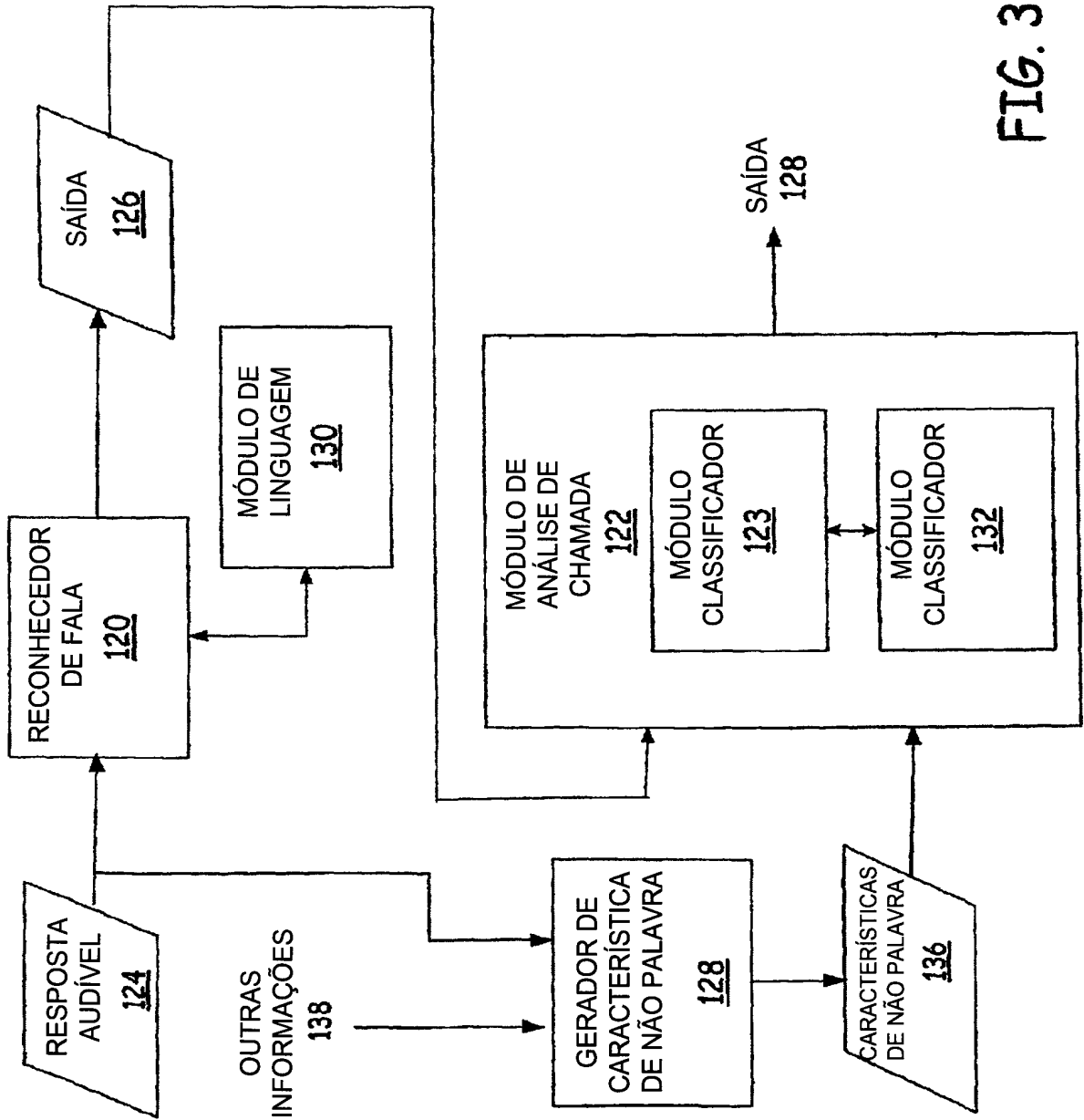


FIG. 3

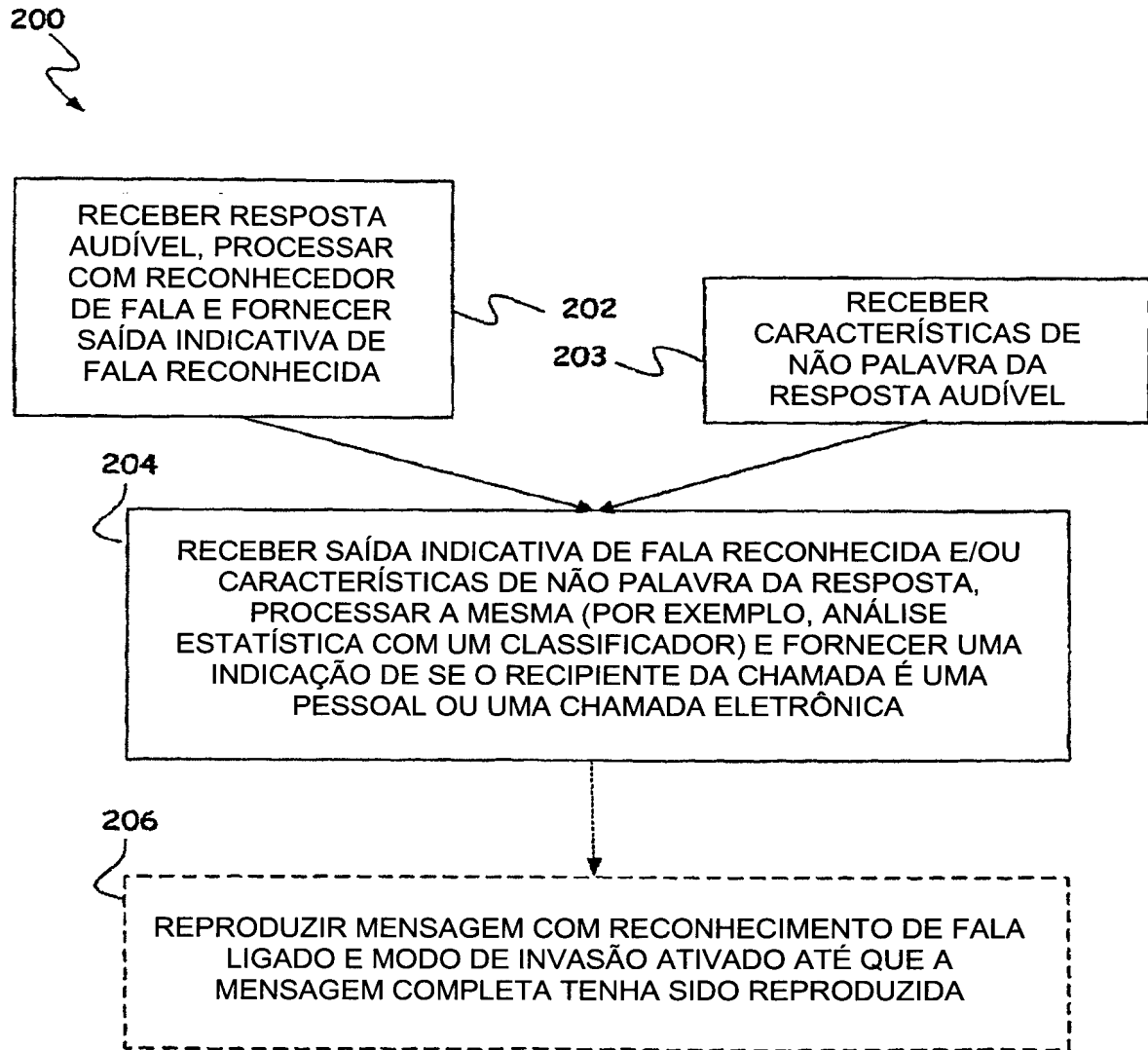


FIG. 4

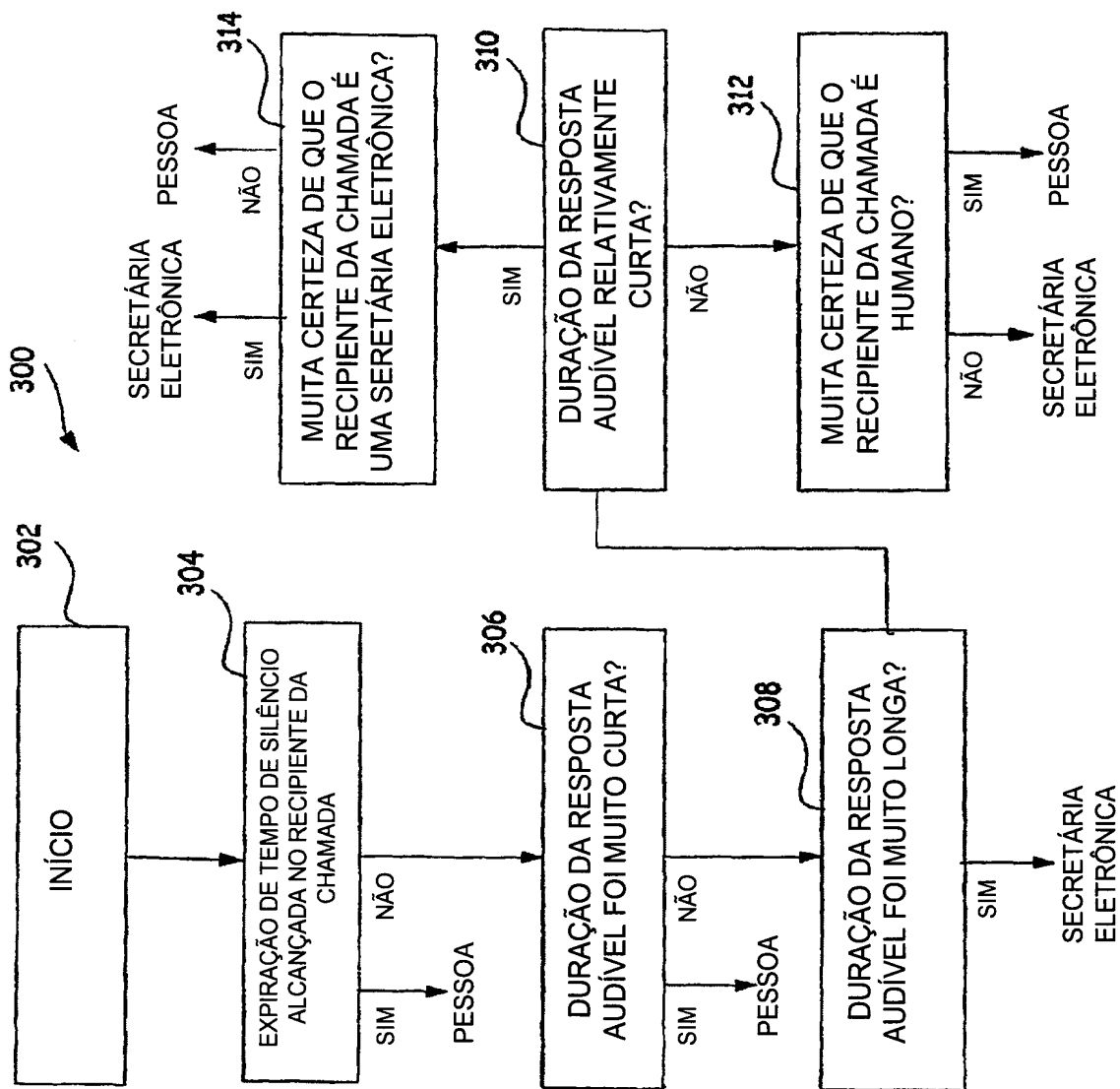


FIG. 5

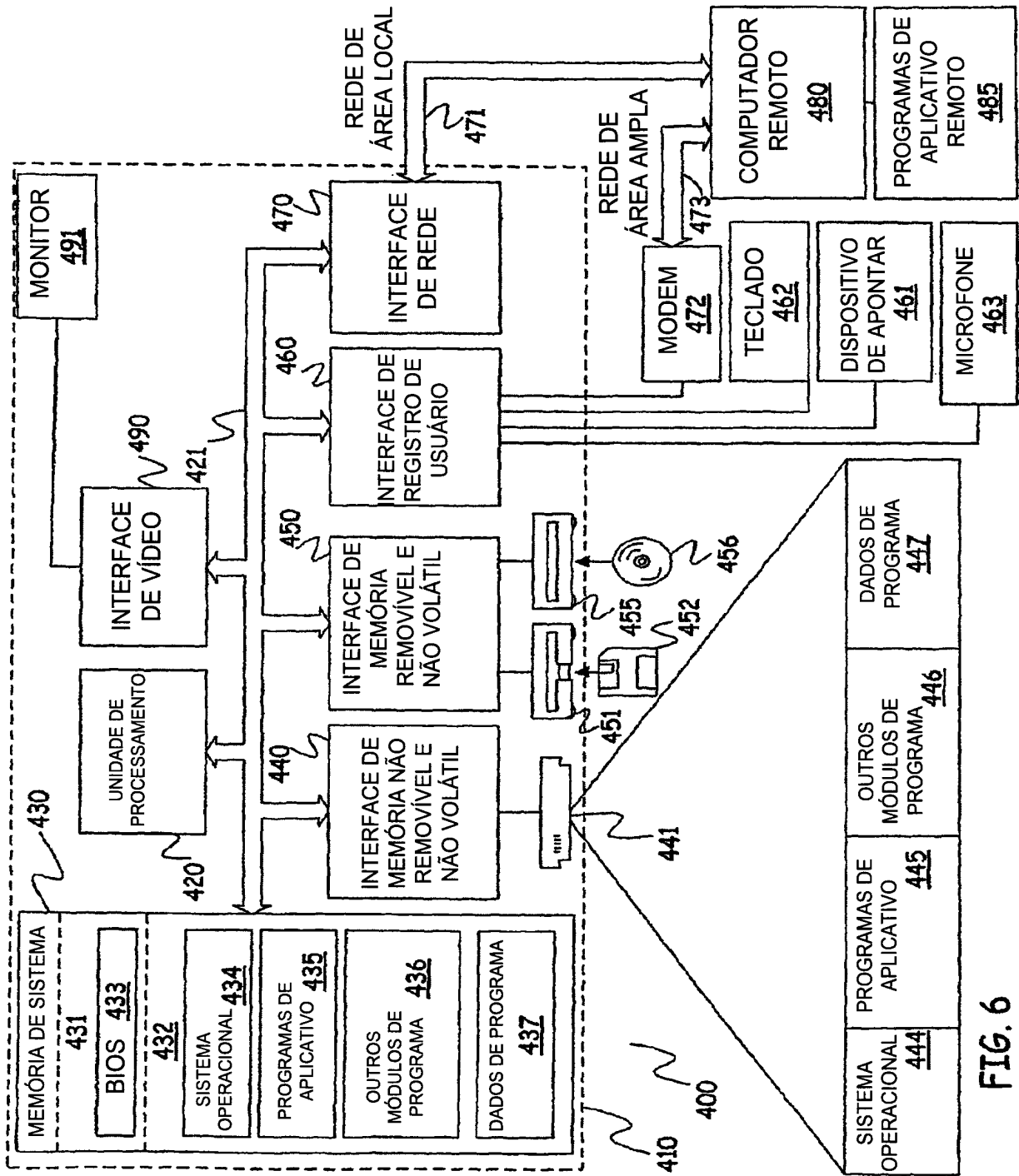


FIG. 6

RESUMO

"DETECÇÃO DE UMA SECRETÁRIA ELETRÔNICA UTILIZANDO RECONHECIMENTO DE FALA"

Um módulo de detecção de secretária eletrônica é utilizado para determinar se um
5 recipiente de uma chamada é uma pessoa ou uma secretária eletrônica. O módulo de detecção de secretária eletrônica inclui um reconhecedor de fala e um módulo de análise de chamada. O reconhecedor de fala recebe uma resposta audível do recipiente da chamada para uma chamada. O reconhecedor de fala processa a resposta audível e fornece uma saída indicativa da fala reconhecida. O módulo de análise de chamada processa a saída do
10 reconhecedor de fala para gerar uma saída indicativa de se o recipiente de chamada é uma pessoa ou uma secretária eletrônica.