Comunicação Audio e Vídeo

Assistentes Pessoais Inteligentes com Reconhecimento de Voz

Afonso Costa Bruno Gonçalves
afonso.costa@tecnico.ulisboa.pt bruno.m.goncalves@tecnico.ulisboa.pt
Instituto Superior Técnico

***Resumo*—No cinema, as menções às assistente pessoais inteligentes são, e desde sempre foram, em filmes de ficção cientifica, obrigatórias. Quem não se lembra do mau feitio de HAL, de *2001: A Space Odissey*? Aproximadamente 40 anos depois disso, nasce Siri, pela Apple, e a revolução das assistente pessoais inteligentes começa. É este o tema tratado neste artigo de divulgação, passando pela evolução histórica, a tecnologia, o mercado e modelo de negocio, enunciando algumas perspetivas futuras e aspetos sociológicos.**

I. Introdução

Durante muitas gerações as expectativas para assistentes pessoais com reconhecimento de voz estavam altas. No cinema, as menções em filmes de ficção científica a este tema eram recorrentes, basta olhar para o exemplo do computador da nave *U.S.S. Enterprise* da saga *Stark Strek*, ou para o grande impacto (talvez menos positivo) de HAL, em *2001: A Space Odissey*, data de 1968. De certa forma, as mentes do público estariam a olear-se, a preparar-se para a revolução e, como em tantas tecnologias que depois surgiram na vida real, os filmes de ficção cientifica elevavam as fasquias.

Efetivamente, com os avanços tecnológicos dos telemóveis e o aparecimento do *smartphone*, este tornou-se o dispositivo ideal para albergar um assistente pessoal com reconhecimento de voz. No entanto, não é só neste tipo de dispositivos que podemos encontrar estes assistentes. Existem soluções para casa ou para o nosso computador pessoal, à medida que assistimos a cada vez mais aparelhos terem uma integração com este tipo de aplicações. Atualmente, a competição entre este tipo de aplicações é feroz, sendo que se alguma empresa apresenta uma nova função, então os seus concorrentes logo se adaptam.

II. Evolução Histórica

Inevitavelmente, antes da existência do primeiro assistente pessoal com reconhecimento de voz, o desafio era criar um dispositivo que reconhece-se as nossas palavras. Pois esse desafio começa, entre as décadas de 50 e 60 [1], quando a Bell Laboratories, em 1952, deu os primeiros (tímidos) passos: começaram pelos números. O AUDREY [2] era totalmente analógico e foi o primeiro dispositivo, documentado, com reconhecimento de voz, capaz de reconhecer números ditos por apenas uma voz.

A IBM estreou-se de seguida com o Shoebox [3]. Este dispositivo conseguia responder a 16 palavras, incluindo os 10 dígitos, tendo a possibilidade de resolver problemas simples de aritmética, como contas de somar.

Apesar destes primeiros sucessos, dispositivos como o AUDREY ainda não eram economicamente atraentes, ocupavam muito espaço, consumiam bastante energia, e por isso, não foram ativamente explorados na altura.

Na década de 70, o reconhecimento de voz sofreu uma grande evolução, com o programa da DARPA, *Speech Understanding Research*. Um dos resultados foi o sistema HARPY, que conseguia perceber 1011 palavras, aproximadamente o vocabulário de um menino de três anos. Neste caso, a revolução passou por terem sido adicionados ao sistema mecanismos de deteção de conjuntos acústicos comuns, eliminando a necessidade de analisar todas as amostras [4]. Esta década ficou também marcada pela introdução de um sistema que já conseguia reconhecer vozes de várias pessoas, por parte da Bell Laboratories.

A introdução do Modelo oculto de Markov [5] elevou o número de palavras reconhecidas das centenas para as milhares, com a possibilidade de reconhecimento de um número ilimitado de palavras. Em consequência, quase todos os sistemas de reconhecimento de voz são hoje baseados neste modelo [6].

O tapete vermelho estava assim a ser lançado. O reconhecimento de voz começou a ser utilizado comercialmente, especialmente para aplicações médicas [1]. Nas casas de algumas pessoas, entrou a Julie, uma boneca que conseguia interagir com quem brincava com ela, através de comandos de voz [7]. Em todo o caso, estes sistemas ainda não eram eficientes, os utilizadores teriam que ditar as palavras e falar pausadamente para serem entendidos.

Em solução a este problema, vem o Dragon NaturallySpeaking em 1997. Este *software*, que hoje é comercializado nas suas versões mais recentes para Windows e macOS [8], conseguia reconhecer as palavras ditas, mesmo que os seus utilizadores falassem naturalmente, detetantdo até 100 palavras por minuto. Ainda nesta década, a BellSouth introduziu o sistema VAL [9], que permitia aos utilizadores ligarem para um computador, através do seu telefone, para receber informações, como noticias, meteorologia, horóscopo, páginas amarelas, entre outras. Talvez nesta fase, este sistema já se aproximava um pouco de um assistente pessoal dos dias de hoje.

Até à revolução introduzida pela Apple, com a Siri, sucessivamente os nossos computadores pessoais vinham integrando a possibilidade da elaboração de operações com comandos de

voz, embora este tipo de utilização não fosse muito popular. A Google também impulsionou muito a evolução desta tecnologia. Repare-se que algo que prendia o escalamento desta tecnologia era a dificuldade das empresas em ter amostras de dados disponíveis e a dificuldade de processa-los eficientemente, de modo a otimizar a perceção daquilo que o utilizador está a dizer, por parte dos sistemas. Ora, em especial, a Google tem uma visão bastante privilegiada deste assunto, visto que a combinação e análise dos dados de milhões de pesquisas no seu motor de busca poderão propulsionar a predição e as sugestões feitas pelos assistentes pessoais inteligentes.

A Google começou então por lançar em 2007 um serviço telefónico, o GOOG-411 [10]. Este serviço gratuito, utilizava reconhecimento de voz para que os utilizadores que ligavam para o serviço, entre outras informações, pudessem obter números de telefone de empresas de uma dada cidade. O objetivo para este serviço, por parte da Google, era claro: obter amostras de voz. O serviço acabou depois por ser descontinuado em 2010. No entanto, em 2011, a Google voltou ao ataque introduzindo o Google Voice Search [11], onde se poderiam invocar pesquisas no motor de busca através da voz, em especial no *smartphone*. Foi precisamente em 2011 que a grande revolução foi feita com a Siri. Esta foi inicialmente desenvolvida pela SRI International [12], resultado de décadas de investigação nas áreas de inteligência artificial. Inicialmente seria integrada em todos os tipos de *smartphone*. No entanto foi adquirida pela Apple em 2010 [13], passando a ser exclusiva dos seus produtos. Atualmente existem inúmeras soluções no mercado que serão referidas de seguida.

III. Tecnologia

A. Arquitetura de Agentes

Os assistentes pessoais inteligentes têm como principal função desempenhar tarefas consoante pedidos do utilizador. Esta arquitetura pode ser vista como um modelo cliente-servidor, em que do lado do cliente temos a aplicação a correr numa máquina e no lado do servidor temos as ferramentas fundamentais para o processamento da informação enviada pelo cliente. Como cada tarefa é realizada dependendo do software a que esta está associada, estas tarefas são divididas por vários agentes[14]. Existe um agente que funciona como regulador, ou neste contexto, o coordenador, um agente de interação com o utilizador e os agentes responsáveis por cada tarefa[14]. Existe ainda um módulo responsável pela interpretação do discurso enunciado pelo utilizador, o processador de fala. O papel do agente de interação com o utilizador prende-se com estabelecer a conexão entre o utilizador e o Coordenador[14]. Quando o utilizador inicia um discurso, os módulo processador de fala atua, gerando um pedido parcialmente analisado. O agente de interação cria então uma mensagem com esta informação para enviar ao agente Coordenador. No fim de todo o processamento necessário para realizar tarefas e responder ao utilizador, o agente de interação responsabiliza-se por converter a resposta do sistema para um formato que possa ser lido pela aplicação[14]. Na figura 1 está descrita esta arquitetura de agentes.

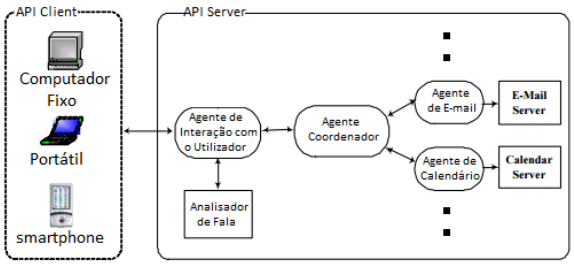

Figura 1. Arquitetura de Agentes

B. Funções do Agente Coordenador

O agente coordenador é responsável por gerir o diálogo estabelecido com o utilizador[14] e coordenar as ações desenvolvidas pelos agentes específicos. Para tal, este agente possui uma série de planos[14] para gerir as tarefas, coordena-las e para manter um diálogo coerente. O agente possui conhecimento sobre o contexto da conversa de modo a conseguir um melhor entendimento dos pedidos do utilizador e a conseguir responder de maneira mais acertada. Com os planos e o conhecimento contextual, o agente coordenador pretende[14]:

- entender os atos comunicativos do utilizador
- perceber as intenções do utilizador
- satisfazer os desejos do utilizador

Percebe-se então que o agente coordenador tem duas funções principais, manter o diálogo e coordenar as atividades dos agentes específicos. No entanto, é importante perceber que estas funções são executadas separadamente. A função de gestão de diálogo é baseada no princípio de reação, isto é, a resposta a pedidos do utilizador, e no principio de proatividade, clarificando os pedidos e notificar o utilizador[15]. De modo a entender qual a intenção do pedido e estabelecer a melhor resposta, o agente coordenador baseia-se no no conhecimento do contexto, que pode ser definido pelos seguintes parâmetros[15]:

- Histórico da conversa
- Lista de conteúdo mencionado, classificando-o segundo o seu grau de ocorrência
- Tarefas realizáveis
- Vocabulário especifico para interpretação de tarefas
- Variáveis de contexto - Ex: Tipo de dispositivo, modalidades

Apesar de num diálogo entre o agente coordenador e o utilizador existirem vários atos de fala específicos, apenas alguns são essenciais para o agente coordenador conseguir entender o tipo de pedido do utilizador. Estes atos são chamados atos conversacionais[14]. Estes atos podem ser executados pelo coordenador, tanto na comunicação com o utilizador como com os agentes específicos. De seguida apresentam-se os vários atos conversacionais[14]:

- Pedido - solicitar a resolução de uma tarefa
- Resposta - descrever o resultado do pedido
- Clarificação - pedir para esclarecer ambiguidades
- Cumprimento - expressar o desejo de iniciar uma conversa
- Confirmação - expressar acordo ou desacordo

- Reconhecimento - mostrar que entendeu o desejo do remetente

Como vimos anteriormente, o agente coordenador baseia-se numa série de planos para conseguir um funcionamento estável. Estes planos podem ser divididos em quatro grupos[14]. Quando chega a mensagem do utilizador, depois de processada pelo módulo de processamento de fala, esta é gerida pelos planos que se encaixam num grupo intitulado Determinação de Atos Conversacionais(grupo CAD). Estes planos são responsáveis por determinar o ato conversacional e por definir quais os tipos de tarefas a executar. De seguida a informação é passada pelo grupo chamado Identificação de Intenções(grupo II). Se estes planos tiverem sucesso, a informação é passada aos planos do grupo Processador de Tarefas(grupo TP), que irá envolver novos pedidos feitos pelo coordenador aos agentes específicos. Se a intenção do utilizador não tenha sido descodificada pelo coordenador, então é realizado um pedido de clarificação ao utilizador. Este pedido de clarificação é realizado pelo último grupo de planos Gerador de Respostas(grupo RG). Na figura 2 ilustra-se um exemplo de funcionamento desta arquitetura de agentes em que o utilizador faz um pedido para ver a última mensagem recebida do João.

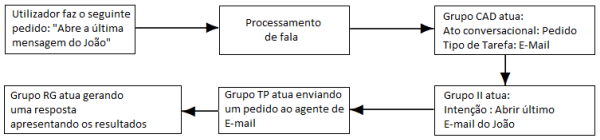

Figura 2. Exemplo de funcionamento

C. Processador de fala

Neste tópico explora-se o funcionamento do módulo de processamento de fala. Antes de mais é necessário entender o formato de entrada deste sistema. A entrada do sistema é um sinal analógico capturado pelo microfone do dispositivo do lado do cliente e enviado em formato digital para o servidor. O principal objetivo deste módulo é construir uma hipótese da sequência de símbolos mais prováveis que correspondem ao sinal de entrada[16]. Sendo E o sinal de entrada e S a sequência de símbolos da hipótese, então o nosso problema pode ser traduzido na seguinte equação[16]:

$$\hat{S} = argmax P(S|E) = argmax \frac{P(E|S)P(S)}{P(E)} \quad (1)$$

A equação 1 baseia-se no teorema de Bayes[16]. Como a probabilidade de um sinal de entrada vai ser igual para todas a sequências de símbolos então a equação pode reduzir-se a[16]:

$$\hat{S} = argmax \frac{P(E|S)P(S)}{P(E)} = argmax P(E|S)P(S) \quad (2)$$

Nesta equação a probabilidade de ocorrer uma sequência de símbolos irá depender apenas da regularidade de ocorrência prévia desse símbolo. A probabilidade de um sinal de entrada ser tal sabendo que originou uma determinada sequência de símbolos pode ser determinada pelos *Hidden Markov Models*[16].

Para determinar estas probabilidades a arquitetura à qual se recorre encontra-se na figura 3[16].

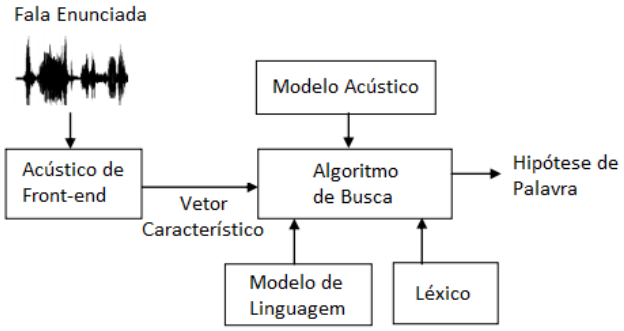

Figura 3. Arquitetura do processador de fala

O modelo de linguagem permite determinar a probabilidade de ocorrência de uma certa sequência de símbolos enquanto que o modelo Acústico permite encontrar a probabilidade de um certa sinal acústico de entrada sabendo a sequência de símbolos[16]. No primeiro bloco desta arquitetura o principal objetivo é determinar vetores característicos[16], que representam o sinal de entrada num formato mais compacto. Este bloco atua em três fases. A primeira trata de fazer uma análise espectral do sinal e gerar características da envolvente espectral de curtos sinais de voz. De seguida, são criados os vetores característicos com a informação dinâmica e estática adquirida. Por último, estes vetores são comprimidos e fortalecidos, de modo a estarem preparados para potenciais erros. As características extraídas devem ser tais que[16]:

- permitam ao sistema diferenciar entre sons relativamente parecidos
- facilitem a criação de modelos acústicos sem necessitar de muita informação adicional
- as suas estatísticas não veriem consideravelmente de utilizador para utilizador

Existem inúmeros métodos para cumprir estes requisitos. O mais utilizado é o *Mel-frequency Cepstral Coefficient* (MFCC). Este método começa por amplificar a energia do sinal nas frequências altas. De seguida parte-se o sinal em ”janelas”. Tendo apenas o sinal que se encontra dentro dessa janela de tempo, aplica-se a transformada de Fourier de modo a obter um representação em fase e amplitude. É importante ter em conta, nesta fase, que o ouvido humano é mais sensível às baixas frequências[17]. Executa-se então uma alteração de escala de frequências para a escala de Mel. A equação para alteração de escala é a seguinte [17]:

$$M(f) = 1125\ ln(1 + f/700) \quad (3)$$

De modo a explorar esta característica auditiva, aplicam-se também filtros triangulares igualmente espaçados nas frequências baixas e logaritmicamente espaçados nas frequências altas[17]. À saída deste processo tem-se o espetro

de *Mel*, que no passo que se segue é sujeito à aplicação do logaritmo, obtendo-se os coeficientes espetrais de *Mel*. Nesta fase os coeficientes gerados são substituídos por coeficientes DCT. Conclui-se assim o processo de criação dos vetores característicos do sinal de entrada.

Passemos agora à descrição da peça fundamental e mais complexa desta arquitetura: o modelo Acústico. A função deste modelo prende-se com determinar representações estatísticas das características geradas[16]. Para tal, baseia-se nos *Hidden Markov Models* (HMM). O HMM é um processo através do qual se pretende determinar qual é o resultado final, tendo apenas informação incompleta. Neste caso, vamos ter vários pedaços de discurso, que vão ser chamados de fonemas. Cada fonema é tratado pelos HMM tendo em conta a linguagem e uma base de dados que fornece palavras e os fonemas que lhe deram origem. Os HMM juntam então vários fonemas de modo a perceber que palavra está a ser construída[18]. O objetivo é encontrar a palavra que maximiza a probabilidade de um determinado conjunto de fonemas.

O modelo de linguagem baseia-se na aprendizagem de contextos. Para tal utiliza textos disponíveis online ou numa base de dados fornecida pela entidade responsável pelo processo de reconhecimento de voz. Ao estudar estes textos, tenta criar um contexto para cada um deles e perceber de que forma as palavras se encaixam num determinado num determinado contexto. Um método frequentemente utilizado para melhorar o desempenho deste processo prende-se com a predição da palavra que o utilizador vai dizer a seguir[16].

IV. Modelo de Negócio

Hoje em dia, todos os grandes nomes da tecnologia têm associados um assistente digital e, se não têm, estão a desenvolver algum. Isto deve-se à facilidade com que estes têm acesso a grandes quantidades de dados e ao poder computacional para os processar e, assim, terem a capacidade de responder a pedidos em tempo real dos utilizadores que utilizem este tipo de aplicações [19].

No entanto, basta uma breve pesquisa na web para perceber que existem inúmeras *startups* [20] que, apesar de não terem o poder computacional, nem o número de clientes que os gigantes da tecnologia têm, conseguem cada vez mais desenvolver produtos deste tipo e pô-los competir com os gigantes. Veja-se por exemplo o caso da Viv Labs [21], uma *startup* que se baseia em inteligência artificial no desenvolvimento de um produto que consegue interagir com todo o tipo de dispositivos. Esta foi co-fundada por alguns criadores da Siri para a criação da Viv, considerada uma extensão mais inovadora e poderosa da Siri, cujo objetivo será de estar integrada em todo o tipo de dispositivos, do *smartphone* ao frigorífico [22]. Em outubro de 2016, a Viv foi adquirida pela Samsung [23]. Neste caso, relativamente ao modelo de negócio, a Viv Labs continuará a operar como uma empresa independente, prestando serviços à Samsung e suas plataformas.

Para o caso de muitas *startups* e empresas que produzem este tipo de *software*, talvez o objetivo passe em um dia vender ou prestar serviços para um gigante da tecnologia. No entanto, o principal modelo de negócio é a comercialização de aplicações, em especial para *mobile*, que oferecem uma assistente pessoal inteligente. Por exemplo, como é o caso da Viv e de outros assistentes pessoais, como o SoundHound Hound [24] e a Braina [25], que são comercializados para os dispositivos móveis e computadores pessoais.

No entanto, existem outras *startups* que, talvez por, como referido, não terem o poder computacional e base de clientes dos gigantes da tecnologia, oferecem assistentes pessoais com foco em alguns tipos de tarefas, como a calendarização de reuniões ou a gestão da caixa de e-mail. No entanto, muitas destas aplicações não tem suporte para o reconhecimento de voz. Ainda assim, são importantes marcos a referir neste trabalho, uma vez que contribuem para a competição no mercado existente das assistentes pessoais com reconhecimento de voz, bem como introduzem novas ideias para a inovação deste tipo de aplicações. Casos das criações destas *startups* são a X.ai, uma assistente pessoal que se foca na calendarização de eventos entre os diferentes envolvidos, ou do EasilyDo, focada na caixa de e-mail.zz

Na maior parte dos casos, estas empresas oferecem uma assistente pessoal com um número limitado de *features* na versão gratuita do seu sistema, sendo esta versão complementada por uma ou mais versões *premium* que os utilizadores têm de pagar. Quanto aos grandes nomes da tecnologia, como a Apple, ou a Microsoft, estes integram os seus assistentes pessoais com reconhecimento de voz nos próprios dispositivos, fazendo parte do pacote. A Google disponibiliza gratuitamente o Google Now, tendo também uma solução paga, o Google Assistant. Em geral, as atualizações do *software* dos *smartphones* trarão melhorias à assistente pessoal, e não são pagas.

No entanto, nem todas as assistentes pessoais existentes no mercado estão como soluções de *software* para dispositivos móveis ou computadores pessoais. Existem soluções para casa deste tipo de dispositivos. Neste caso, vende-se um dispositivo que permite gerir com comandos de voz os outros aparelhos da casa, tocar música ou, por exemplo, encomendar uma pizza. Assim, o que se reflete neste tipo de dispositivos é que permitem em muito a integração com IoT, tendo sempre associado uma aplicação para *smartphone*. Em geral, as atualizações de *software* deste dispositivo também são gratuitas, sendo que a maneira com que as empresas ganham dinheiro será na venda inicial do produto.

V. Mercado Atual

Como percebido pelo modelo de negócio, o mercado atual é muito dinâmico no que toca a a soluções de assistentes pessoais inteligentes, apresentando soluções que já vêm integradas no *smartphone* e computadores pessoais, soluções a nível de software que podemos adquirir, ou soluções para casa [26].

Para o primeiro caso, as soluções mais conhecidas já integradas em dispositivos, são a Siri (2011), da Apple, a Cortana (2014) da Microsoft e o Google Now (2012), da Google, representados na figura 4. Estas três soluções encontram-se respetivamente integradas nos *smartphones* ou computadores pessoais das respetivas marcas, sendo possivelmente as mais

utilizadas no mercado. No caso dos *smartphones* o Google Now é uma aplicação mais ampla a nível do número de dispositivos com os quais é compatível, podendo ser utilizada tanto em Android como iOS. Nesse sentido, não consegue uma integração tão boa quanto a Siri, e a Cortana, que são construidas para um número mais limitado de *smartphones* **??**. Abrir uma aplicação, por exemplo, é possível utilizando a Cortana e a Siri mas com o Google Now pode não ser possível em certos dispositivos. Quanto aos computadores pessoais, a Cortana vem integrada no Windows 10 e a Siri passou a ser integrada na ultima versão do macOS, o macOS Sierra. Recentemente, após a recente aquisição da Samsung da assistente Viv, existe em eminente a possível integração desta nos próximos *smartphones* da Samsung, adicionando mais uma solução para o mercado [28].

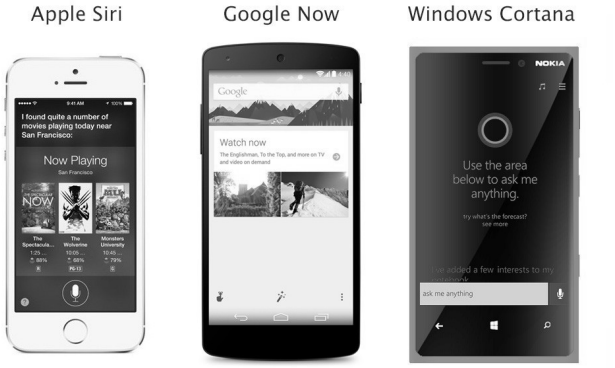

Figura 4. Os três grandes assistentes pessoais: Siri, Google Now e Cortana.

No entanto, em alternativa a estas aplicações, existem outras que podem ser adquiridas. O Hound, da SoundHound é um exemplo que pode ser adquirido gratuitamente para o *smartphone* através da loja de aplicações. A Braina é outro assistente pessoal que pode ser adquirido para o Windows, existindo uma versão gratuita e uma versão PRO, com subscrição de 3 anos, que pode chegar aos $117.

A nível de soluções para casa, a grande competição está entre o Amazon Echo e o Google Home, que estão representados na figura 5. O Amazon Echo está disponível por $179.99 e o Google Home por $129.00. Ambos permitem a interação com outros dispositivos de casa, como também têm a componente normal de uma assistente pessoal, sendo que a assistente pessoal do primeiro é a Alexa e do segundo o Google Assistant. Conseguem, portanto, realizar todo o tipo de pedidos nesse sentido, como encomendar uma pizza, por exemplo, ou chamar um Uber.

Numa era em que cada vez mais o poder da comunidade se mostra impulsionador das tecnologias, o Mycroft mostra-se uma solução *Open Source* e *Open Hardware*. Este é mais um assistente pessoal para casa, tal como o Amazon Echo, cujo o seu *software* e *hardware* está inteiramente documentado *online*. Isto permite a que, em conjunto, todos os utilizadores possam desenvolver aplicações para este dispositivo e partilhalas com os outros membros da comunidade.

Figura 5. Os assistentes pessoais para casa: Amazon Echo e Google Home.

VI. Aspetos Legais e Segurança

Ao contrário do assistente pessoal inteligente KIT, de *Knight Rider*, os assistentes pessoais de hoje ainda não nos protegem contra bandidos. Mas poderão estar a ser, pelo contrário, os próprios bandidos?

Reparemos que estes sistemas são desenhados para fazer as nossas vidas mais fáceis, com a possibilidade de fazerem algumas das tarefas mais chatas por nós, basta pedirmos. No entanto, para o fazerem, necessitamos de dar acesso a estas aplicações a uma quantidade bastante significativa de informação pessoal. Estas informações passam pelo acesso ao nosso calendário, ao nosso email, contactos telefónicos, entre muitos outros. Para isso, muitas vezes teremos que fornecer as nossas palavras passe a estes sistemas, o que poderá abrir portas para muito mais [29].

Além disso, repare-se como estes sistemas trabalham. Se queremos, por exemplo, pesquisar algo na *internet*, dizemos à assistente pessoal para o fazer, esta grava a nossa voz pedindo a desejada pesquisa e envia-a para um *data center* onde a aplicação remota realmente processa o pedido. Ou seja, o nosso pedido, a nossa gravação do mesmo, que deverá ser privada, vai estar guardada num sítio que provavelmente não sabemos onde é, num país diferente do nosso, onde algumas das leis serão totalmente diferentes. Neste sentido, tudo o que perguntamos ao nosso assistente pessoal inteligente é guardado e processado completamente fora do nosso controlo [29].

As questões que eventualmente se podem colocar é se poderemos ou não confiar neste tipo de aplicações com as nossas informações. Quanto à questão destas aplicações terem acesso a grande parte dos nossos dados, é claro que a decisão de confiar numa empresa com os nossos dados será nossa, não só neste caso em particular, mas como em muitos outros, desde o primeiro momento em que utilizamos a *internet*. Por exemplo, ao utilizarmos um *chat online*, como o Facebook Messenger, ou o Whatsapp, estamos a confiar que as empresas que fazem a gestão destas aplicações não irão espiar as nossas conversas, apesar de poderem ter o poder de o fazer. Ou, noutros casos, quando nos registamos noutro serviço e confiamos que essa dada empresa não utilizará os nossos dados pessoais de registo para outros fins.

Existem exemplos da recolha de dados feita por este tipo de aplicações. É o caso da assistente pessoal da Microsoft, a Cortana, incluída nas mais recentes versões do Windows 10. Nas políticas de privacidade deste serviço [30], lê-se: *"A Cortana foi criada para ser a sua verdadeira assistente pessoal,*

fornecendo-lhe sugestões e alertas pessoais relevantes. Para tal, a Cortana tem de compreender determinados dados sobre si, tais como os seus interesses, localizações e preferências". Ou seja, tal como dito, a Cortana necessita de dados pessoais do utilizador para que a sua utilização plena. Ainda assim, neste caso, a Microsoft dá a possibilidade do utilizador decidir quais os dados que quer partilhar com a Cortana, que vão desde os dados de localização, até ao calendário ou mesmo o histórico de pesquisa. Neste caso, as funcionalidades da Cortana ficarão limitadas ao tipo de acessos que o utilizador lhe dá. Além disso, como descrito na política de privacidade anterior, sempre que a Cortana necessita de um novo acesso, esta pedirá primeiramente ao utilizador.

Ainda em relação ao caso da Cortana, caso não desativemos uma outra opção existente, os dados recolhidos podem ser utilizados para a Microsoft atribuir um perfil publicitário ao utilizador e posteriormente condicionar a apresentação de publicidade mediante cada utilizador. Medidas como esta costumam ser muito polémicas entre a comunidade utilizadores.

Em muitos outros casos, para as próprias empresas poderem melhorar o seu *software*, estas optam por incluir nas políticas de privacidade das suas ferramentas termos que lhes permitem recolher dados ao longo da utilização do utilizador da mesma, por via à realização de melhorias existentes.

No fundo, ao utilizarmos estes serviços, estamos a comprometer-nos com o fornecimento de dados pessoais para que a assistente pessoal possa ser mais eficiente nas suas operações.

VII. Aspetos Sociológicos

Ao estudar uma tecnologia é altamente importante perceber qual o impacto desta a nível social. É então relevante uma discussão sobre quais as vantagens e desvantagens do uso desta tecnologia. Nesta secção, debatemos em primeiro lugar as vantagens e depois as desvantagens.

Esta tecnologia permite a um utilizador acesso a informação de uma forma menos trabalhosa, ou seja, faz um pedido e o assistente pessoal trata desse pedido pelo utilizador. Sejamos sinceros, a nível de procura e tratamento de dados um processador é bastante mais eficiente e faz com que as tarefas se realizem mais rapidamente, poupando tempo ao utilizador [32]. Com este tipo de tecnologia a comunicação entre as pessoas pode ser fomentada. Enquanto que um utilizador para enviar o e-mail tem de preencher o campo destinatário, assunto e escrever a mensagem, um assistente pessoal adquire estas informações todas de uma vez do utilizador e cria o e-mail a ser enviado, podendo até criar um assunto percebendo o contexto da mensagem[32]. Também dentro de uma organização, o papel do assistente pessoal inteligente pode ser importante[33]. Nesta ótica, seriam vários utilizadores a comunicar com um ou mais assistentes pessoais. Estes, então, falam entre si, permitindo uma maior eficiência na organização interna. Esta possibilidade de aceder a informação tão rapidamente faz com que o utilizador deixe de fazer pesquisas. Isto resulta em a pesquisa ser baseada apenas no conteúdo e não no formato apelativo que este possa ter [34]. Ora, os avanços que foram feitos na elaboração de *websites* estão em risco de terem cada vez menos relevância à medida que esta nova tecnologia se apodera da sociedade. Não é no entanto claro se este fator é uma vantagem ou uma desvantagem.

No entanto, nem tudo o que esta tecnologia traz tem impacto positivo na sociedade. De seguida apresentam-se então alguns aspetos negativos da introdução desta tecnologia nas nossas vidas. Estes assistentes pessoais permitem ao utilizador realizar tarefas que ele próprio iria realizar ou então iria contratar alguém para as fazer. Percebe-se então que este tipo de tecnologia faz com que alguns empregos deixem de fazer falta. Isto tem um elevado impacto na sociedade pois pode fazer aumentar o nível de desemprego[32]. Apesar de terem de existir pessoas a desenvolver o *software* e *hardware* necessário, este fator tem pouca relevância no impacto geral no desemprego. Outro ponto fundamental que deve ser tido em conta é o facto de que este mecanismo de fácil acesso à informação e sensação de conforto pode provocar nas pessoas um afastamento da vida real e falta de comunicação direta com outras pessoas. Estes fatores podem levar à quebra de relações e ao aumento da solidão [32]. Outro impacto importante prende-se com a evolução humana. Sabe-se que a espécie humana tem evoluído consoante os seus requisitos para sobreviver. Ora, esta tecnologia faz com que o utilizador não precise de usar capacidades cognitivas avançadas para a realização de funções. Percebe-se então que algumas capacidades cognitivas vão desaparecer ao longo do tempo dado que a necessidade de as usar deixa de ser preponderante [35].

VIII. Perspetivas Futuras

Para quem assistiu ao filme *Her* [36] de 2013, sabe bem que no que toca às assistentes pessoais inteligentes, o limite é inexistente. Neste filme, passado em Los Angels, algures no futuro, a tecnologia de assistentes pessoais inteligentes é bastante avançada. Os *operating systems* (OS), como lhes chamam, têm a habilidade de aprender e crescer psicologicamente a um ritmo elevado, ultrapassando a largos passos a capacidade humana. Isso, psicologicamente. O que significa que estas aplicações, neste filme, conseguem de alguma forma simular sentimentos e emoções. Numa quebra de paradigma, começa a ser aceitável socialmente a existência de relações amorosas entre as pessoas e os seus OS's. O impressionante, neste caso, não é as pessoas apaixonarem-se por estas máquinas, mas sim o contrário, os OS apaixonarem-se, o que mostra um extremo avanço tecnológico. Quão distantes estamos nós de uma sociedade destas?

Apesar de ser apenas um filme, *Her* ilustra bem as inúmeras possibilidades existentes neste tipo de aplicações, o poder de processamento que estas eventualmente terão e, sobretudo, a controvérsia lançada a nível social. Mas afinal, para um futuro muito mais próximo do que *Her*, por quais avanços tecnológicos estas aplicações poderão passar?

Ora, como já todos nos temos apercebido, a barreira dos assistentes pessoais inteligentes serem uma aplicação limitada ao nosso *smartphone* está a ser ultrapassada. Cada vez é mais habitual utilizar a nossa linguagem natural para falar com a nossa *smart TV* de casa, para procurar conteúdos de um certo realizador ou onde um certo ator entra. Por exemplo,

na mais recente *box* da operadora NOS permite a execução de comandos de voz para fazer este tipo de operações. No entanto, as aplicações das assistentes pessoais inteligentes não ficam por aqui. Os carros, por exemplo, também têm sido algumas das vítimas na integração deste tipo de sistemas.

Quanto ao próprio sistema dos assistentes pessoais, é difícil perceber qual será a próxima revolução. A existência deste tipo de sistemas é relativamente recente, não nos permitindo perceber qual será o próximo *big thing* [37]. Para já, as próximas versões destas aplicações serão seguramente melhores, mais inteligentes e mais rápidas do que as anteriores, mas ainda dentro do mesmo tipo de conceito, não fugindo ao tipo de produto da versão anterior. No entanto, como aconteceu ao longo de toda a evolução tecnológica, inevitavelmente as assistentes pessoais poderão dar lugar a novos produtos totalmente diferentes, e à quebra de paradigmas. Quem sabe, talvez mais parecidos com Samantha, um OS do filme *Her*.

Neste sentido, dentro do conceito atual de assistentes pessoais inteligentes, ainda há muito para aprender em alguns aspetos. Por exemplo, no que diz respeito ao quão bem a assistente pessoal interpreta o que queremos, e à relevância dos conteúdos que procura por nós. Muitas vezes estas aplicações não têm a habilidade de inferir o que necessitamos e quais as nossas intenções, baseando-se apenas naquilo que lhes perguntamos. Ou seja, os resultados apresentados por estas são baseados apenas no conteúdo por elas encontrado e não no formato apelativo que este possa ter. Não conseguem fazer a distinção, que nós conseguimos fazer, entre o próprio conteúdo e a sua qualidade.

Este facto leva a que as próprias respostas das assistentes pessoais sejam, por vezes, demasiado vagas. As assistentes pessoais inteligentes são produtos com o objetivo de chegar a um mercado amplo, sem distinção do tipo de pessoas que eventualmente as pode usar. Neste sentido, a informação que conseguem adquirir mediante os nossos pedidos também é ampla. Além disso, à medida que vamos explorando o nível especifico do tipo de tarefas que queremos que estas façam, também teremos que abrir novas portas à quantidade de informação a que permitimos acesso. Assim sendo, talvez para estas assistentes pessoais serem cada vez mais inteligentes, seja necessário que estas se especifiquem na elaboração de apenas algumas tarefas.

Para o futuro, a evolução talvez passe pelo aperfeiçoamento das assistentes pessoais nestes tópicos mencionados. Os esforços na investigação em disciplinas como a inteligência artificial, a aprendizagem automática ou *data mining* poderão ser fundamentais para a evolução dos assistentes pessoais. Neste momento, a capacidade de processamento ou o reconhecimento de voz já não são os grandes desafios. Com as evoluções nestas áreas, os assistentes pessoais poderão ter um comportamento mais pro-ativo no que diz respeito às informações passadas ao utilizador. Isto é, atualmente, a maior parte das ações feitas por estas aplicações são tomadas após os comandos de voz fornecidos pelo utilizador. No entanto, o objetivo para o futuro passará pelas aplicações executarem automaticamente ações, baseando-se em informações que vão adquirindo sem a ação direta do utilizador ou em informação encontrada por elas (automaticamente) *online*.

IX. Conclusão

Neste artigo, apresenta-se a evolução dos assistentes pessoais inteligentes desde o seu aparecimento, focando-se em revelar uma arquitetura baseada em agentes que realizam tarefas em substituição do utilizador. Ora, este fato levanta algumas questões sociais e legais que foram discutidas de modo a perceber qual poderá ser o melhor futuro para este tipo de tecnologia. Aborda-se também os vários produtos presentes no mercado de modo a suscitar no leitor um sentido crítico sobre cada uma delas.

X. Referências

[1] Melanie Pinola
Speech Recognition Through the Decades: How We Ended Up With Siri
Disponível em: http://www.pcworld.com/article/243060/speech_recognition_through_the_decades_how_we_ended_up_with_siri.html
Acesso em: 5 de dezembro de 2016.

[2] Roberto Pieraccini
From AUDREY to Siri. Is speech recognition a solved problem?
Disponível em: http://www.icsi.berkeley.edu/pubs/speech/audreytosiri12.pdf
Acesso em: 5 de dezembro de 2016.

[3] IBM Archives: IBM Shoebox
Disponível em: https://www-03.ibm.com/ibm/history/exhibits/specialprod1/specialprod1_7.html
Acesso em: 5 de dezembro de 2016.

[4] Lowerre, B. T.
The Harpy speech recognition system
Disponível em: http://adsabs.harvard.edu/abs/1976PhDT.....81L
Acesso em: 5 de dezembro de 2016.

[5] Hidden Markov Model
Disponível em: https://en.wikipedia.org/wiki/Hidden_Markov_model
Acesso em: 5 de dezembro de 2016.

[6] Mark Gales and Steve Young
The Application of Hidden Markov Models in Speech Recognition
Disponível em: http://mi.eng.cam.ac.uk/~mjfg/mjfg_NOW.pdf
Acesso em: 5 de dezembro de 2016.

[7] Worlds of Wonder "Julie"(an interactive talking doll)
Disponível em: http://www.dollinfo.com/wowjulie.htm
Acesso em: 5 de dezembro de 2016.

[8] Dragon Speech Recognition Software
Disponível em: http://www.nuance.com/dragon/index.htm
Acesso em: 5 de dezembro de 2016.

[9] ALTech's SpeechWorks System Brings Speech-Enabled Yellow Pages to BellSouth Customers
Disponível em: http://www.prnewswire.com/news-releases/altechs-speechworks-system-brings-speech-enabledyellow-pages-to-bellsouth-customers-76147252.html
Acesso em: 5 de dezembro de 2016.

[10] Wikipedia: GOOG-411
Disponível em: https://en.wikipedia.org/wiki/GOOG-411
Acesso em: 5 de dezembro de 2016.

[11] Wikipedia: Google Voice Search
Disponível em: https://pt.wikipedia.org/wiki/Google_Voice_Search
Acesso em: 5 de dezembro de 2016.

[12] Sri International - Siri
Disponível em: https://www.sri.com/engage/ventures/siri
Acesso em: 5 de dezembro de 2016.

[13] Erick Schonfeld
Silicon Valley Buzz: Apple Paid More Than $200 Million For Siri To Get Into Mobile Search
Disponível em: https://techcrunch.com/2010/04/28/apple-siri-200-million/
Acesso em: 5 de dezembro de 2016.

[14] Wayne Wobcke
Disponível em: https://pdfs.semanticscholar.org/719f/812a5a6648cf44af7e02f515e8d292861f17.pdf
Acesso em: 5 de dezembro de 2016.

[15] UNSW
Disponível em: http://www.cse.unsw.edu.au/wobcke/spa-seminar.pdf
Acesso em: 5 de dezembro de 2016.

[16] Karpagavalli
Disponível em: http://www.sersc.org/journals/IJSIP/vol9_no4/34.pdf
Acesso em: 5 de dezembro de 2016.

[17] Crypto
Disponível em: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/
Acesso em: 5 de dezembro de 2016.

[18] CMU
Disponível em: http://www.cs.cmu.edu/roni/11761/Presentations/hmm-for-asr-whw.pdf
Acesso em: 5 de dezembro de 2016.

[19] Alex Brokaw and Ben Popper
Here's how startups are outsmarting Siri and Alexa
Disponível em: http://www.theverge.com/2016/7/8/12110036/artificial-intelligence-startups-open-source-api-google-viv-easilydo-hound
Acesso em: 5 de dezembro de 2016.

[20] Angel.co
Intelligent Assistants Startups Disponível em: https://angel.co/intelligent-assistants
Acesso em: 5 de dezembro de 2016.

[21] Viv Labs
Disponível em: http://viv.ai/
Acesso em: 5 de dezembro de 2016.

[22] Zoë Corbyn
Meet Viv: the AI that wants to read your mind and run your life
Disponível em: https://www.theguardian.com/technology/2016/jan/31/viv-artificial-intelligence-wants-to-run-your-life-siri-personal-assistants
Acesso em: 5 de dezembro de 2016.

[23] Matthew Panzarino
Samsung acquires Viv, a next-gen AI assistant built by the creators of Apple's Siri
Disponível em: https://techcrunch.com/2016/10/05/samsung-acquires-viv-a-next-gen-ai-assistant-built-by-creators-of-apples-siri/
Acesso em: 5 de dezembro de 2016.

[24] SoundHound Hound
Disponível em: http://www.soundhound.com/hound
Acesso em: 5 de dezembro de 2016.

[25] Braina
Disponível em: https://www.brainasoft.com/braina/
Acesso em: 5 de dezembro de 2016.

[26] Wikipedia - Intelligent Personal Assistants Disponível em: https://en.wikipedia.org/wiki/Intelligent_personal_assistant
Acesso em: 5 de dezembro de 2016.

[27] Predictive Analytics Today
Top 19 Intelligent Personal Assistants or Automated Personal Assistants
Disponível em: http://www.predictiveanalyticstoday.com/top-intelligent-personal-assistants-automated-personal-assistants/
Acesso em: 5 de dezembro de 2016.

[28] Tim Hardwick
Viv on Samsung Galaxy S8
Disponível em: http://www.macrumors.com/2016/11/07/samsung-galaxy-s8-debut-viv-creators-siri/
Acesso em: 5 de dezembro de 2016.

[29] Gavin Kenny
I Know Everything About You! The Rise of the Intelligent Personal Assistant
Disponível em: https://securityintelligence.com/i-know-everything-about-you-the-rise-of-the-intelligent-personal-assistant/
Acesso em: 5 de dezembro de 2016.

[30] Políticas de privacidade Cortana
Disponível em: https://privacy.microsoft.com/pt-PT/windows-10-cortana-and-privacy
Acesso em: 5 de dezembro de 2016.

[31] The Windows 10 privacy issues you should know about Disponível em: http://thenextweb.com/microsoft/2015/07/29/wind-nos/
Acesso em: 5 de dezembro de 2016.

[32] Karehka Ramey
Disponível em: http://www.useoftechnology.com/advantages-disadvantages-information-technology/
Acesso em: 5 de dezembro de 2016.

[33] Steven Okamoto
Disponível em: http://www.cs.cmu.edu/pscerri/papers/Okamoto_SPA_Book_Chapter.pdf
Acesso em: 5 de dezembro de 2016.

[34] Tom-Anthony
Disponível em: https://moz.com/blog/intelligent-personal-assistants-replace-websites
Acesso em: 5 de dezembro de 2016.

[35] Mike Fekety
Disponível em: https://www.linkedin.com/pulse/pros-cons-artificial-intelligence-mike-fekety
Acesso em: 5 de dezembro de 2016.

[36] Roger Ebert - Film Summary
Her Movie Review (2013)
Disponível em: http://www.rogerebert.com/reviews/her-2013
Acesso em: 5 de dezembro de 2016.

[37] Peter Sweeney
Siri's Descendants - How intelligent assistants will evolve
Disponível em: https://medium.com/adventures-in-consumer-technology/siris-descendants-fd36df040918
Acesso em: 5 de dezembro de 2016.

XI. Autores

Bruno Gonçalves
Nascido a 10 de Novembro de 1995. É estudante de Engenharia Eletrotécnica e de Computadores, nas áreas de computadores e telecomunicações. Escolheu este curso porque desde pequeno tem fascínio por computadores, eletrónica e tudo o que era relacionado.

Afonso Costa
Nascido a 18 de Abril de 1995. Estudante de Engenharia Eletrotécnica e de Computadores, na área de computadores e secundária telecomunicações. Durante o seu percurso académico desenvolveu vários projetos relacionados com circuitos eletrónicos e algoritmos computacionais.