



Marcela Pedroso Maués

**Um olhar sobre os assistentes virtuais
personificados e a voz como interface**

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós-graduação em Design da PUC-Rio como requisito parcial para obtenção do grau de Mestre em Design.

Orientadora: Claudia Mont'alvão

Rio de Janeiro
Maio de 2019



Marcela Pedroso Maues

**Um olhar sobre os assistentes virtuais
personificados e a voz como interface**

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre pelo Programa de
Pós-Graduação em Design da PUC-Rio. Aprovada
pela Comissão Examinadora abaixo.

Claudia Renata Mont'Alvão Bastos Rodrigues

Orientador

Departamento de Artes & Design - PUC-Rio

Marcelo Fernandes Pereira

Departamento de Artes & Design - PUC-Rio

Berenice Santos Gonçalves

Universidade Federal de Santa Catarina - UFSC

Rio de Janeiro, 06 de Maio de 2019

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização do autor, do orientador e da universidade.

Marcela Pedroso Maués

Graduou-se em Desenho Industrial (PUC-Rio, Pontifícia Universidade Católica do Rio de Janeiro) em 2015. Trabalha como UX Designer / Arquiteta da Informação desenvolvendo produtos digitais.

Ficha Catalográfica

Maués, Marcela Pedroso

Um olhar sobre os assistentes virtuais personificados e a voz como interface / Marcela Pedroso Maués ; orientadora: Cláudia Mont'Alvão. – 2019.

116 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Artes e Design, 2019.

Inclui bibliografia

1. Artes e Design – Teses. 2. Assistentes virtuais. 3. Interface de voz. 4. Experiência de uso. 5. Design de interação. 6. Interação humano-computador. I. Mont'Alvão, Cláudia. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Artes e Design. III. Título.

CDD: 700

Para meus pais, Mario e Titi, para meu irmão, João Paulo, e para todos os
amigos e professores que me acompanharam durante essa jornada.

Agradecimentos

À minha orientadora Professora Claudia Renata Mont’Alvão Bastos Rodrigues pelo apoio, incentivo e parceria para realização desse trabalho.

À Capes e à PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Aos meus pais, Mario Jorge e Titi, pela educação, ensinamentos e carinho.

Aos meus amigos e ao meu irmão, por toda palavra amiga, apoio e compreensão.

Às minhas gatas, Mika e Nena, pela companhia constante durante os dias e noites de trabalho.

Resumo

Maués, Marcela Pedroso; Rodrigues, Claudia Renata Mont'Alvão Bastos. **Um olhar sobre os Assistentes Virtuais Personificados e a voz como interface**. Rio de Janeiro, 2019. 116p. Dissertação de Mestrado - Departamento de Artes & Design, Pontifícia Universidade Católica do Rio de Janeiro.

Com cada vez mais objetos interconectados com capacidade de processamento próprios, a computação gradativamente se desvincula do computador *desktop*, e se torna mais favorável a novos formatos de computadores e novas interfaces. Os computadores que conversam e as interfaces de voz já foram largamente explorados em filmes e livros de ficção como "*Ela*", "*2001: Uma Odisseia no Espaço*" e "*Eu, robô*", dentre outros. No entanto, não é apenas na ficção que a voz como interface aparece: sendo cada vez mais comum nos dias atuais, ela é forma de interação em vários dispositivos como computadores *desktop*, *smartphones*, *smartwatches*, *smartTvs* e *videogames*. Essa interface é também adotada em assistentes virtuais que auxiliam seus usuários em toda sorte de funções do dia-a-dia como organização de agenda, anotações, previsão do tempo, despertadores e controle dos chamados objetos inteligentes. O foco dessa pesquisa foi a relação dos usuários com essas interfaces de voz e com os assistentes virtuais, e como está se dando a recepção dos mesmos: como está acontecendo sua adoção, quais as opiniões e expectativas do usuário, e que problemas encontra? A premissa da pesquisa é que a personificação, ou seja, a humanização através da adoção de uma personalidade e persona, bem como o uso de interfaces conversacionais nos assistentes virtuais facilita a sua aceitação e torna a opinião sobre os mesmos positiva. Para responder essa questão, foi realizada uma revisão bibliográfica e uma pesquisa de campo em duas etapas: a aplicação de um questionário online e a realização de grupos de foco. O questionário foi aplicado com o intuito de recrutar participantes para o grupo de

foco, além de investigar a adoção e o uso de assistentes de voz por pessoas que moram no Rio de Janeiro. A aplicação do grupo de foco visou levantar a aceitação e as opiniões desses usuários sobre esses Assistentes Virtuais Personificados que usam a voz como interface, e mapear suas dificuldades com a tecnologia e suas expectativas acerca da mesma para o futuro. O resultado mostrou que a personalidade desses assistentes realmente facilita a sua adoção e tem efeitos positivos sobre a percepção do usuário, porém as interfaces de voz ainda geram desconforto quando usadas em ambiente público. As maiores preocupações levantadas pelos usuários em relação aos assistentes de voz não foram técnicas ou de usabilidade, mas reflexões sobre o impacto dos assistentes na sociedade e na vida das pessoas.

Palavras-chave:

Assistentes virtuais, Interface de voz, Experiência de uso, Design de interação, Interação humano-computador, Ergodesign

Abstract

Maués, Marcela Pedroso; Rodrigues, Claudia Renata Mont'Alvão Bastos. **A view on personified virtual assistants and voice as an user interface.** Rio de Janeiro, 2019. 116p. Dissertação de Mestrado - Departamento de Artes & Design, Pontifícia Universidade Católica do Rio de Janeiro.

As there are increasingly more interconnected objects with processing capacity, computing gradually disconnects from the desktop computer as it becomes more open to new computer formats and new interfaces. Talking computers and voice interfaces have been extensively explored in films and fiction books such as "*Her*", "*2001: A Space Odyssey*" and "*I, Robot*", among others. However, it is not only in science fiction that the voice user interface appears. It became more common these days as a form of interaction in several devices like desktop computers, smartphones, smartwatches, smartTVs and videogames. This interface is also adopted in virtual assistants that help its users in all sorts of day-to-day functions such as scheduling, annotations, weather forecasting, alarm clocks, and smart objects control, and some of the virtual assistants have been developed with a humanized approach, adopting a name and a personality. The focus of this research was the relationship of users with these voice interfaces and personified virtual assistants, and how they are being received: how is their adoption taking place, what are the opinions and expectations of its users, and what problems do they encounter? The hypothesis of this research is that the personification and the adoption of conversational interfaces in virtual assistants facilitates its acceptance and makes the opinions about it more positive. To answer this question, a bibliographical review and a field research were carried out in two stages with an online form and focus groups. The questionnaire was applied in order to investigate the adoption and use of voice assistants by people living in Rio de Janeiro, and to recruit participants to the focus group. The focus group aimed to explore the acceptance and opinions of

these users about the personified virtual assistants that use the voice as interface, and to map their difficulties and their expectations for the future of technology. The result shows that the personality of these assistants facilitates their adoption and has positive effects on user perception, but voice interfaces still generate discomfort when used in the public environment. The major concerns users faced while using voice assistants were not technical or usability problems but worries about the impact of the assistants on society and people's lives.

Keywords:

Virtual assistants, Voice user interface, User experience, Interaction design, Human-computer interaction, Ergodesign

Sumário

1. Introdução	15
1.1 Tema, problema e objeto de pesquisa	18
1.2 Relevância	18
1.3 Premissa e questões norteadoras	19
1.4 Objetivos geral e específicos	19
2. O design de interação na era da computação ubíqua	21
2.1 O que é design de interação?	21
2.2 O atual paradigma da computação	24
2.3 Os desafios desse contexto para o designer	27
2.3.1 Introdução de novas experiências com sucesso	28
2.3.2 Projetar para a adaptação	30
2.3.3 Experiência contínua em diferentes dispositivos	31
2.3.4 Interconexão de dispositivos	32
2.3.5 Discutir a complexidade e os limites do uso da tecnologia	33
2.4 As interfaces para a computação ubíqua	35
2.5 Resumo do capítulo	40
3. Os Assistentes Virtuais Personificados e a fala	42
3.1 A voz e a fala como formas de interação	46
3.2 Interfaces conversacionais	51
3.3 Resumo do capítulo	55
4. O residente do Rio de Janeiro e os Assistentes Virtuais Personificados que usam a voz como interface	57
4.1 Questionário	57
4.1.1 Características gerais dos respondentes	59

4.1.2 Características do perfil dos respondentes	62
4.1.3 Sobre o uso dos assistentes pelo perfil selecionado	66
4.1.4 Conclusões sobre o questionário	70
4.2 Grupo de foco	73
4.2.1 Sobre a realização do grupo de foco	75
4.2.2 Os problemas tecnológicos	77
4.2.3 Usabilidade e eficiência em assistentes	79
4.2.4 Os momentos e usos dos assistentes	82
4.2.5 Acessibilidade	84
4.2.6 Os assistentes no ambiente conectado	85
4.2.7 A customização e a personalização dos assistentes	86
4.2.8 A humanização e o assistente personificado	88
4.2.9 A voz como forma de interação com AVPs	90
4.2.10 A voz dos assistentes	91
4.2.11 A questão social e os impactos na vida das pessoas	93
5. Considerações finais e desdobramentos	101
6. Referências bibliográficas	104
7. Apêndices	106
7.1. Questionário online	106
7.2. Termo de consentimento	111
7.3. Roteiro do grupo de foco	112

Lista de figuras

Figura 1 - Dispositivos utilizados para acessar a internet no Brasil	16
Figura 2 - Newton, PDA da Apple	26
Figura 3 - Site do Supermercado Pão de Açúcar utilizando os recursos de breadcrumbs e menu	37
Figura 4 - Google Assistente exibindo informações visuais complementares	44
Figura 5 - Uso de imagem para personificar um assistente	45
Figura 6 - Tecnologias envolvidas na interface conversacional	54

Lista de gráficos

Gráfico 1 - Distribuição de respondentes por local de moradia	59
Gráfico 2 - Distribuição etária dos respondentes	60
Gráfico 3 - Escolaridade dos respondentes	61
Gráfico 4 - Distribuição por gênero	62
Gráfico 5 - Distribuição etária entre respondentes do Rio de Janeiro maiores de idade	63
Gráfico 6 - Escolaridade dos respondentes maiores de idade do Rio de Janeiro	63
Gráfico 7 - Declaração de gênero dos respondentes do Rio de Janeiro com mais de 18 anos	64
Gráfico 8 - Dispositivos utilizados para acessar a internet pelos respondentes maiores de idade do Rio de Janeiro	65
Gráfico 9 - Uso recente de assistentes virtuais por respondentes maiores de idade residentes no Rio de Janeiro	66
Gráfico 10 - Assistentes mais usados pelos respondentes do perfil	67
Gráfico 11 - Meios de acesso aos assistentes	68
Gráfico 12 - Principais objetivos de uso dos assistentes	69
Gráfico 13 - Experiência e adoção dos assistentes	70
Gráfico 14 - Adoção dos assistentes entre que usou ou não nos 3 meses anteriores	70
Gráfico 15 - Grau de adoção dos residentes do Rio de Janeiro maiores de idade dividido por assistente	71
Gráfico 16 - Avaliação da experiência de uso por assistente	72

Samantha: Well... Right when you asked me if I had a name I thought "yeah, he's right, I do need a name". But I wanted to pick a good one, so I read a book called "How to Name Your Baby", and out of a hundred and eighty thousand names that's the one I liked the best.

Theodore: Wait, you read a whole book in the second that I asked what your name was?

Samantha: In two one hundredths of a second actually.

Her (Spike Jonze, 2013)

1

Introdução

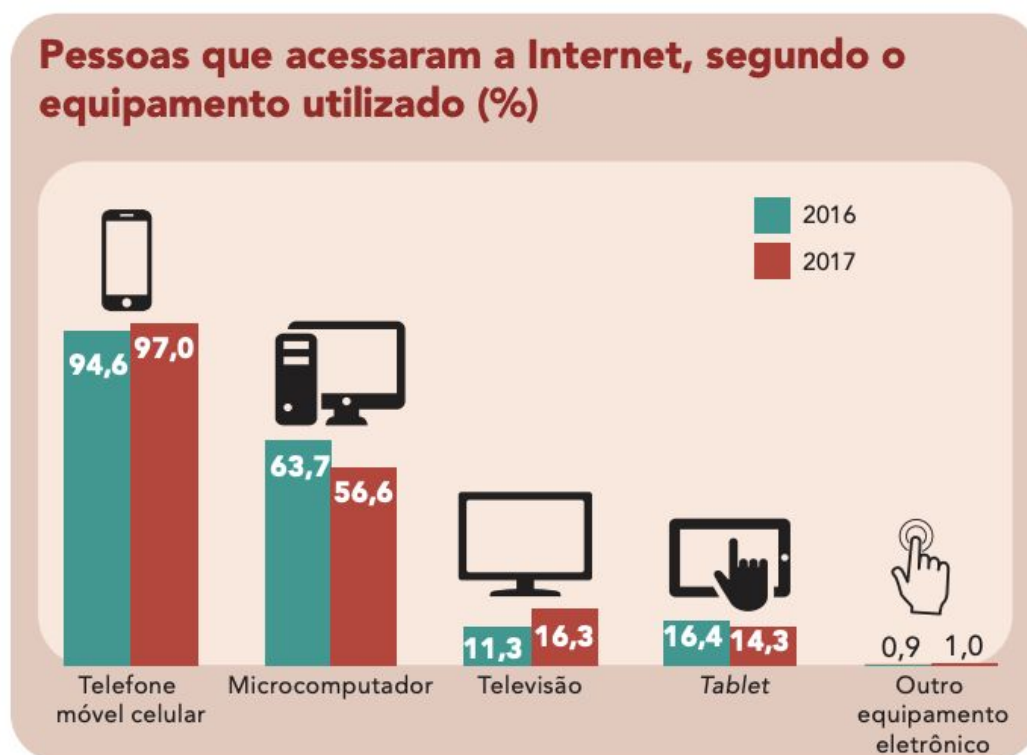
Desde o surgimento dos computadores pessoais e sua posterior conexão a nível mundial através da *internet*, a computação vem evoluindo, e junto vem também nossa forma de interagir com ela. Essa evolução abre caminho para novos *inputs* e *outputs* que servem à interação com os novos computadores.

Atualmente, existem interfaces de voz, interfaces gestuais, realidade aumentada, realidade virtual, realidade mista e muitas outras tecnologias. Essas formas de interação vêm se tornando cada vez mais comuns à medida que novas tecnologias diminuem de custo e se tornam mais populares, tomando conta de televisores, celulares e videogames.

Os assistentes virtuais são sistemas que podem performar tarefas ou realizar serviços para seu usuário. Atualmente, assistentes que auxiliam em funções do dia-a-dia, como organização de agenda, anotações, previsão do tempo, ajuste de despertadores, controle de objetos inteligentes¹ e outras, estão disponíveis nos grandes sistemas operacionais de celular, *iOs* da Apple e *Android* do Google. Por estarem disponíveis em português em celulares de baixo custo com acesso à *internet*, os assistentes virtuais e as interfaces de voz aumentam seu alcance no Brasil. Segundo a Pesquisa Nacional por Amostra de Domicílios contínua 2016/2017² (PNAD Contínua) do IBGE, dos 69,9% dos brasileiros que acessam a *internet*, 97% usam a *internet* em telefones móveis, conforme mostra a figura 1.

¹ Os objetos inteligentes são dispositivos computacionais embutidos em objetos do dia-a-dia que, conectados entre si e dotados de microprocessadores e sensores, são capazes de identificar contextos e responder a dados cenários como que portadores de uma inteligência.

² Dados mais atualizados até a data de publicação.



Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional por Amostra de Domicílios Contínua 2016-2017.

FIGURA 1 - Dispositivos utilizados para acessar a internet no Brasil

Os assistentes virtuais têm a voz e o texto como principais métodos de entrada de dados (*input*). Para essa pesquisa, vamos considerar os seguintes tipos de interface: **comando** de voz ou texto, como em assistentes virtuais de *call center* e *telemarketing*, e **interfaces conversacionais**, como na Google Assistente, Siri, Cortana e Alexa.

Esses assistentes podem, também, ser humanizados: adotar uma persona com voz própria, opinião, humor e personalidade. Dessa forma, se aproximam dos assistentes ou secretários reais, que existem fora do mundo virtual, e com quem é possível criar laços. Esse é o caso da Siri da *Apple* e da Google Assistente da *Google*³.

³ Siri e Google Assistente são assistentes pessoais que estão nos telefones celulares de sistema operacional *Android* e *iOS* mais recentes, também sendo encontrado em outros dispositivos, como computadores *desktop*. Os celulares mais antigos não trazem os assistentes

Essa pesquisa irá tratar desse grupo de assistentes virtuais humanizados e que usam a interface conversacional como forma de interação, os quais estaremos chamando de **Assistentes Virtuais Personificados** ou **AVPs**.

Para os *designers* de interação, profissionais encarregados de projetar as interfaces entre os humanos e os computadores ou artefatos digitais, é mais importante entender como se estabelece a relação dos usuários com esses novos objetos e sistemas, que entender a engenharia que permite a evolução dessas tecnologias. Estudar como as pessoas adotam um produto, como usam-no, quais suas necessidades, os problemas que encontram, o que gostam e não gostam, e se conseguem realizar as tarefas pretendidas, é atividade constante conforme os ambientes se tornam mais complexos.

Apesar de termos literatura e estudos acerca desse tipo de interface, os dados costumam tratar do uso de assistentes fora do Brasil, em países de língua inglesa, padrão na maioria dos sistemas. É difícil encontrar material que trate o público brasileiro, e os *designers* precisam contar com dados de um usuário diferente para projetar para esse tipo de interação. Por esse motivo, faz-se necessário estudar a recepção da voz como interface pelo brasileiro e entender se e como se encaixa nas necessidades cotidianas e particularidades desse público.

Sabendo que o sucesso de um produto não se dá somente pelas características objetivas do mesmo, essa dissertação se propôs a estudar a relação dos usuários com os Assistentes Virtuais Personificados. Observando a aceitação e a opinião sobre esses sistemas e sobre a voz como forma de interação com eles, a pesquisa se propôs a examinar a premissa de que a personificação e o uso de interfaces conversacionais nos assistentes virtuais facilita a sua aceitação e torna a opinião sobre os mesmos positiva.

embarcados no sistema, sendo necessária uma atualização de *software* ou a troca do aparelho. A voz é utilizada como uma interface conversacional onde se cria um diálogo com o assistente, ela é *input* e *output* do sistema.

1.1 Tema, problema e objeto de pesquisa

O tema dessa pesquisa é o olhar sobre os Assistentes Virtuais Personificados que usam a voz como forma de interação, pretendendo responder o seguinte problema: hoje, como se dá a aceitação e como é a opinião acerca dos assistentes que usam a voz como interface?

O objeto da pesquisa é, portanto, a relação dos usuários com os Assistentes Virtuais Personificados que usam a voz como forma de interação.

1.2 Relevância

A pesquisa é relevante para a área do *design* de interação, visto a necessidade do *designer* de, ao desenvolver um projeto, compreender seu público e o uso das interfaces que serão exploradas. Entendendo os cenários de uso nos quais dado tipo de interação tem vantagens, é possível fazer escolhas de projeto mais assertivas. Da mesma forma, ao conhecer as desvantagens, é possível trabalhar os aspectos negativos da interface visando sua evolução e favorecendo o desenvolvimento de novas formas de interação.

Além disso, com a maioria das pesquisas tratando dos assistentes e do público falante de língua inglesa, padrão para esses sistemas, os *designers* que projetam para o público brasileiro precisam usar de dados sobre um usuário de perfil diferente. O estudo dessas interfaces se torna ainda mais pertinente dado o cenário atual da computação, que ao não precisar mais, necessariamente, das GUIs (*Graphical User Interface*, ou interface gráfica), cria a oportunidade de novas formas de interação com diferentes *inputs* e *outputs*.

1.3 Premissa e questões norteadoras

A premissa dessa pesquisa é que a personificação e o uso de interfaces conversacionais nos assistentes virtuais facilita a sua aceitação e torna a opinião sobre os mesmos positiva.

Também serão norteadoras para o trabalho as seguintes questões:

- Qual o impacto da fala na experiência de uso de Assistentes Virtuais Personificados?
- O uso da voz, no lugar de texto escrito como acontece em *chatbots*, é entendido como benéfico pelo seu usuário?
- Os estudos levantados durante a revisão bibliográfica refletem as particularidades e casos de uso do público estudado? Quais as diferenças e pontos de atenção quando se trata desse público?
- Esse público usa ou pretende usar essa interface no futuro?
- Como está se dando aceitação e quais são as opiniões acerca dos Assistentes Virtuais Personificados?

1.4 Objetivos geral e específicos

Essa pesquisa tem como principal objetivo interpretar e apontar, do ponto de vista do design de interação, os desafios e oportunidades de uso da voz como interface em Assistentes Virtuais Personificados ao se tratar do público estudado.

São considerados os seguintes objetivos específicos:

- Apontar o contexto dos Assistentes Virtuais Personificados que usam a voz como interface, passando pela computação ubíqua e nossa relação com a fala.
- Descrever os resultados encontrados em estudos já realizados sobre assistentes e interfaces de voz.
- Em pesquisa com o público, determinar a adoção dos AVPs e as opiniões sobre os mesmos.
- Identificar como o público estudado aceita e opina sobre a voz como forma de interação com Assistentes Virtuais Personificados

2

O design de interação na era da computação ubíqua

Ao realizar um projeto, o designer precisa estudar os pontos de contato entre seu usuário e o produto. O projeto se torna mais complexo conforme os pontos de contato e as possíveis interações com o mesmo se diversificam.

Na computação ubíqua, as formas de interação evoluíram muito e as possibilidades se ampliaram. Nesse contexto, surgem novas interfaces e novas formas de lidar com os objetos e com o mundo ao redor, o que traz para o designer de interação um novo desafio, mas também novas oportunidades.

Os computadores que conversam e as interfaces de voz, como um todo, já foram largamente explorados na ficção como, por exemplo, no filme *Her* (JONZE, 2014). Nesse filme, a voz é a principal forma de interação entre o personagem Theodore e seu sistema operacional, Samantha. Essa interface também aparece na ficção de *2001: Uma Odisséia no Espaço* (KUBRICK, 1968), *Ex_Machina* (GARLAND, 2015), *Passageiros* (TYLDUM, 2016) e *O Espaço entre nós* (CHELSON, 2017), "Eu, robô" (ASIMOV, 1950), "O Guia do Mochileiro das Galáxias" (ADAMS, 1979), entre outros. No entanto, não é apenas na ficção que a voz como interface aparece, sendo cada vez mais comum atualmente em vários dispositivos como computadores *desktop*, *smartphones*, *smartwatches*, *smart tvs* e *videogames*.

2.1 O que é design de interação?

Bill Moggridge (2007) defende que as disciplinas do design são definidas pelas diferentes limitações que suas soluções podem ter. Por exemplo, se temos um problema que será resolvido através de uma impressão bidimensional,

tipografia e gravuras, a disciplina de design que se insere nesse tipo de limitação é o design gráfico.

A computação possibilitou o crescimento de três grupos de profissionais: designers de interface humano-computador (*HCI designers*), desenvolvedores e especialistas em fatores humanos. O primeiro grupo criava os *softwares* que eram usados nos computadores *mainframes*, nos microcomputadores e nos computadores pessoais. O segundo grupo programava e desenvolvia esses *softwares* com uma visão técnica e com foco em performance. Por fim, o terceiro grupo, de especialistas em fatores humanos, com conhecimentos sobre psicologia, testava e avaliava esses projetos que já estavam prototipados. Dessa forma, havia melhorias incrementais, porém não era encorajada nenhuma inovação mais radical. (Moggridge, 2007)

Moggridge (2007) percebeu, então, na primeira metade dos anos 80, a oportunidade para uma nova disciplina de design. Essa disciplina se dedicaria à elaboração de soluções criativas e atrativas para um mundo virtual, de forma que poderiam ser projetadas não apenas as formas, mas os comportamentos esperados, as animações e os sons. Ele imaginava essa disciplina como o equivalente ao design industrial porém para *software*, e não objetos tridimensionais. Para ele, esse profissional iria estudar as necessidades e desejos dos usuários de um dado produto ou serviço, e se empenharia para criar projetos que trouxessem tanto satisfação quanto prazer estético.

Saffer (2010) resume interação como uma troca entre duas entidades. Geralmente, uma troca de informação, mas que também pode ser uma troca com produtos ou serviços. Por tanto, o designer de interação, projeta para a interação que ocorre entre pessoas, máquinas e sistemas em uma variedade de combinações.

O design de interação é uma disciplina difícil de definir. Em parte, pela sua origem interdisciplinar que vem do design industrial, da comunicação, dos fatores humanos e da interação humano-computador. Porém, é também complicado por

tratar do comportamento, que é um aspecto muito mais difícil de observar e entender que a aparência. (Saffer, 2010)

Antes de realizar uma ação, o indivíduo passa por um processo mental, que não é passível de ser observado. Sua percepção sobre um produto, o entendimento de como funciona, a intenção que tinha quando decidiu como agir - todos esses fazem parte de um processo que ocorre dentro da pessoa. Essa complexidade faz com que o design de interação não seja o projeto da interação em si, mas o projeto para a possibilidade da interação.

Saffer (2010) defende que existem três grandes escolas de pensamento quando se trata de definir essa disciplina de design: o ponto de vista centrado na **tecnologia**, o centrado no **comportamento** e o ponto de vista do design de **interação social**.

Para ponto de vista centrado na tecnologia, o design de interação tem a função de tornar os aparatos tecnológicos digitais úteis, fáceis de serem utilizados e prazerosos de se usar. Os designers de interação, então, pegam o que foi criado por engenheiros e programadores de forma crua, e moldam em produtos que as pessoas de fato gostem de usar. (Saffer, 2010)

Já no ponto de vista centrado no comportamento, design de interação trata de "definir o comportamento de artefatos, ambientes e sistemas (como por exemplo, produtos)" (SAFFER, 2009, p.5 apud FORLIZZI, Jodi e REIMANN, Robert, 1999). Assim, o foco dessa visão é na funcionalidade, no comportamento dos produtos, e no *feedback* (resposta ou retorno) baseado no que os usuários desse artefato estão fazendo.

O último e mais amplo ponto de vista é o do design de interação social. Essa definição coloca o design de interação como uma prática inerentemente social, que gira em torno de facilitar a comunicação entre humanos e produtos. Esse jeito de comunicar pode tomar muitas formas, podendo ser **um para um**, como em uma ligação telefônica, **um para vários**, como em um *site* ou *blog*, ou **vários**

para vários, como é o caso de um mercado de ações. A tecnologia se torna irrelevante, de modo que qualquer objeto ou dispositivo pode fazer uma conexão entre pessoas. (Saffer, 2010)

2.2 O atual paradigma da computação

A computação evoluiu muito desde os computadores com interação por cartão perfurado até os computadores atuais que, por vezes, se escondem em um relógio ou uma pequena jóia. Essa evolução continua acontecendo e é muito favorável à criação de novas tecnologias e interfaces.

Os computadores *mainframes*, que foram majoritários muitos anos atrás, eram gigantes, necessitavam de diversas pessoas para manuseá-los, e possibilitavam apenas algumas funções muito específicas em cada máquina. Com o conceito de Internet das Coisas⁴ se tornando mais comum a cada dia, os objetos cotidianos passam a ter maior poder de processamento, se interconectar, e a ideia da computação gradativamente se desvincula do computador *desktop*. Essa evolução se insere no contexto da computação ubíqua.

Weiser (1993) explica que a computação ubíqua tem como objetivo aprimorar o uso dos computadores, tornando um grande número deles acessível no meio físico, mas ainda assim torná-los "invisíveis". Com esse termo, Weiser não tenta passar a ideia de uma computação que não pode ser vista, mas de uma computação que passa despercebida no cotidiano de seus usuários.

Segundo Weiser (1993), os computadores estão isolados da vida cotidiana e, ao mesmo tempo, isolam as pessoas que os usam da situação a sua volta. Por conta disso, eles falham na sua incumbência de facilitar uma tarefa. Ao invés de

⁴ J. Gubbi et al. (GUBBI et al., 2013) explica que, apesar do termo Internet das Coisas ter evoluído desde sua primeira aparição em 1999, a ideia original defendida permanece: Computadores que percebem a informação no ambiente sem a necessidade da ação ou da intervenção humana. Hoje em dia, se considera não apenas o captar da informação ao redor e o interagir com o mundo físico, mas também usar os padrões atuais da *internet* para prover serviços de transferência de informações e dados, análise, aplicações e comunicação.

se tornarem a ferramenta através da qual uma tarefa é realizada (o foco estaria na tarefa e, por tanto, a ferramenta desapareceria da percepção), o computador muitas vezes permanece como o foco da atenção. É importante notar que Weiser se refere ao contexto da forma que o observa em sua época, e que a computação evoluiu muito desde então. Apesar do tempo passado, Weiser ainda é referência para falar sobre computação ubíqua e muitas das suas reflexões permanecem relevantes hoje em dia.

"A maioria dos computadores que farão parte desse ambiente virtual incorporado será invisível, tanto enquanto fato como enquanto metáfora. Computadores em interruptores de luz, termostatos, aparelhos de som e fornos já ajudam a ativar o mundo. Essas máquinas e outras serão interconectadas em uma rede ubíqua." (WEISER, 1991, tradução nossa)

O *Newton* da *Apple* (figura 2), lançado em 1993, era um assistente virtual acessado por meio de um equipamento portátil PDA (*Personal Digital Assistant*, assistente pessoal digital), cuja tela era manipulada através de uma caneta específica, e não foi um grande sucesso. Apesar de muito diferente da *Siri*, atual assistente virtual da *Apple*, o *Newton* pode ser considerado um precursor. Ele se aproxima da ideia de assistente que temos hoje por ser um serviço que pretende auxiliar na realização de algumas tarefas de maneira contextualizada. O *Newton*, ao ter um nome de gente, também se aproximava da ideia de humanização de assistentes que temos hoje, assim como a *Siri*.

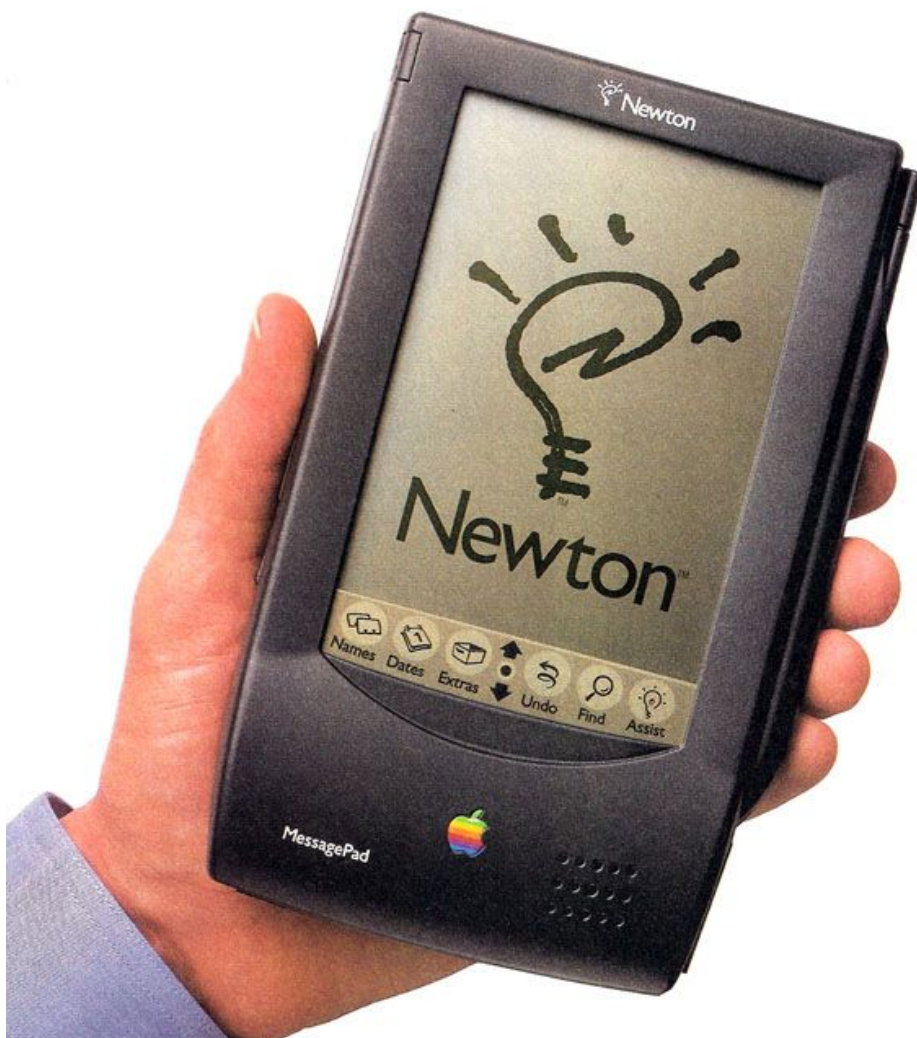


FIGURA 2 - *Newton*, PDA da *Apple*

Os dispositivos conectados que temos hoje, como smartphones, smartwatches, *smartspeakers* e *smarTVs*, entre outros, ajudam a compôr um cenário que se encaminha para o momento de computação ubíqua que Weiser previa. A computação está dividida em diversos artefatos no ambiente, com funções mais específicas ao contexto de uso, e que nem sempre dependem de uma tela.

Os *softwares* atuais também evoluem para se adaptar a múltiplos aparelhos em uma interação mais fluida. As empresas que desenvolvem os grandes sistemas operacionais, como *Google*, *Windows* e *Apple*, trabalham para adaptar seus

sistemas aos diferentes formatos de tela (ou sua ausência), bem como integrar todos em uma rede ubíqua.

Mesmo em aplicativos menores, como é o caso de reprodutores de música, vemos o encaminhamento para o cenário de interconexão de dispositivos. Atualmente, é possível que uma pessoa comece a ouvir uma música em um computador *desktop* em casa e, ao sair para a rua, continue escutando a mesma música em um celular, do mesmo ponto em que o áudio parou.

2.3 Os desafios desse contexto para o designer

No início da computação pessoal, as interfaces de computador usavam de metáforas com o mundo físico para explicar e trazer familiaridade para o novo. No atual estado da computação, o digital já é mais familiar e permite a criação de seus próprios padrões de interação mais abstratos.

Nesse cenário híbrido que mescla o tangível (*hardware*), o intangível (*software*) e serviços, não há mais a necessidade constante de telas e reforços visuais. Surge, então, a necessidade de novas metáforas não-visuais para explicar funcionalidades e trazer entendimento.

Com a abundância de dados capturados sobre as pessoas e seu ambiente, a informação se torna um novo material a ser trabalhado pelo designer. Dar forma, som, cor, bem como esconder ou mostrar informação no mundo físico e digital de acordo com o usuário, seu contexto e a tangibilidade da interface é um dos novos desafios desse cenário complexo.

Mike Kuniavsky (2010) levanta 5 desafios que ele acredita que serão comumente encontrados ao se pensar experiências para esse novo cenário. Usaremos esses 5 pontos para pautar os tópicos de discussão do capítulo.

2.3.1 Introdução de novas experiências com sucesso

Como tempo e dinheiro são limitados, tudo que é novo deve justificar o gasto desses dois recursos. Toda nova forma de interação compete com outra já existente e, por tanto, precisa trazer benefícios claros para ser adotada e justificar seu desenvolvimento. (Kuniavsky, 2010)

O sucesso de algo novo depende da sua aceitação e, para tanto, da sua funcionalidade, facilidade de uso, mas também do aspecto emocional que evoca. O nosso comportamento muitas vezes é resultado de um processo subconsciente, ou seja, bastante coisa é determinada antes de atingir o nível da consciência.

Norman (2008) defende que os objetos de design trabalham três dimensões: visceral, comportamental e reflexiva; dimensões essas que tratam, ao mesmo tempo, de aspectos relativos à emoção e à cognição. Apesar desses geralmente serem trabalhados como aspectos antagônicos, Norman defende que ambos estão presentes nos objetos do dia-a-dia e dialogam; não estando, por tanto, em oposição, e não devendo ser tratados separadamente.

"Nos primeiros anos do computador pessoal, monitores com telas coloridas não existiam. A maioria das telas tinha imagem em preto-e-branco. Sem dúvida, o primeiro computador Apple, o Apple II, podia exibir imagens em cores, mas para jogos: qualquer trabalho sério era feito em preto-e-branco; geralmente texto branco em fundo preto. No início da década de 1980, quando as telas coloridas foram apresentadas ao mundo dos computadores pessoais, tive dificuldade de compreender sua atração. Naquele tempo, a cor era usada principalmente para marcar textos ou acrescentar uma decoração supérflua à tela. Do ponto de vista cognitivo, a cor não acrescentava nenhum valor que o contraste e o sombreamento não pudessem oferecer. Porém, as empresas insistiam em comprar monitores coloridos a um custo mais alto, a despeito de não terem qualquer justificativa científica. Evidentemente, a cor estava satisfazendo alguma necessidade, mas não uma que pudéssemos aferir.

Peguei emprestado um monitor colorido para ver qual era o motivo de toda aquela discussão. Logo me convenci de que minha avaliação original estava correta: a cor não acrescentava nenhum valor que se pudesse discernir no trabalho cotidiano. Contudo, eu me recusei a abrir mão do monitor colorido. Meu raciocínio me dizia que a cor não era importante, mas minha reação emocional me dizia o contrário." (NORMAN, 2008, p.29)

As emoções auxiliam na tomada de decisões: enquanto emoções positivas facilitam o pensamento criativo e a análise de muitas alternativas, as emoções negativas, como medo e ansiedade, colaboram com o foco, com o estreitamento dos processos de raciocínio. Além de defender que emoção e cognição andam juntas, muitas vezes uma influenciando a outra, Norman afirma que o lado emocional do design muitas vezes impacta mais no sucesso de um produto que seus aspectos práticos. (Norman, 2008)

Dan Saffer (2010) corrobora com essa visão dizendo que, apesar de a emoção ser comumente vista como um impedimento para a lógica e para a tomada correta de decisões, no design essa visão precisa ser alterada. Os produtos que não possuem um componente emocional ficam sem vida e não se conectam com as pessoas; a emoção deve, por tanto, ser pensada e incluída nas decisões de *design*.

Além do desafio de conquistar o usuário na adoção de um novo produto, Kuniavsky (2010) levanta também um segundo obstáculo para uma introdução bem sucedida de uma nova experiência: deve-se estar atento ao tempo. Um produto pode fracassar caso demore no desenvolvimento e já esteja ultrapassado no momento de lançamento para o mercado.

Saffer (2010) também expõem a dependência do *design* de interação de uma época e de um contexto. Ele exemplifica citando o navegador mosaico criado por Marc Andreessen em 1994 que, eventualmente, evoluiu para o navegador *Netscape*. Esse navegador foi uma excelente proposta de *design* de interação, tendo introduzido diversos conceitos importantes até hoje para o campo, como o botão de voltar. No entanto, é pouco provável que alguém nos dias de hoje instale esse navegador em seu computador, afinal ele servia bem a um dado contexto e época, que não são os atuais.

A aceitação de um novo produto, por tanto, não é meramente um processo racional que leva em consideração apenas os aspectos práticos de sua utilização. O fator emocional, esse muito mais subjetivo, é componente crucial em se tratar da adoção e do gosto por um produto. Também é de extrema importância o momento

de lançamento, visto que a velocidade das mudanças na tecnologia torna produtos obsoletos cada vez mais rápido.

2.3.2 Projetar para a adaptação

Segundo Kuniavsky (2010), o aumento do poder computacional do *chip* comum fez com que o preço dos dispositivos e do processamento distribuído diminuísse, levando a computação para milhões de pessoas. Em contrapartida, essa diminuição no preço também favoreceu ciclos de lançamento menores com incrementos marginais, que deixam de lado a flexibilidade e a longevidade do produto. Um das adversidades, por tanto, é que mais pessoas passam a ter produtos que são substituídos rapidamente, pois possuem uma vida maior que seu período de popularidade.

Os produtos deveriam ser pensados visando criar experiências que ao mesmo tempo endereçam os desejos e as necessidades imediatas das pessoas, como também permitam a sua adaptação no futuro. Essa abordagem permite que, mesmo que algumas partes não sejam mais funcionais, o produto em si ainda possa ser útil. (Kuniavsky, 2010)

Por vezes, a obsolescência do *software* é pensada. Isso ocorre, por exemplo, quando vemos a possibilidade de transferência de conta e arquivos de um aparelho celular para o outro, na hora de realizar uma troca. Ainda assim, essa solução ainda é muito problemática e frágil, pois muitos desses programas permitem transferência apenas dentro de um mesmo sistema operacional ou para uma mesma categoria de dispositivo. A adaptação contínua visando a utilidade e a segurança é essencial para o design genuinamente centrado no usuário. (Kuniavsky, 2010)

2.3.3 Experiência contínua em diferentes dispositivos

Uma experiência única que perpassa os diferentes dispositivos é uma promessa antiga das redes conectadas. Cada dispositivo, com suas características próprias, permitiria uma interação diferente com um mesmo documento ou *software*. Esse se adaptaria às características do aparelho onde está sendo exibido, extendendo sua experiência de uso. (Kuniavsky, 2010)

Conforme os dispositivos se multiplicam, novas formas de interação também surgem, e o desafio de manter a consistência fica mais complexo. De acordo com Kuniavsky (2010), a consistência é um aspecto fundamental para a criação de uma experiência única, e a manutenção dessa consistência pode ocorrer de diversas formas, de acordo com as necessidades do sistema:

- **Consistência operacional** na maneira de executar as mesmas tarefas em diferentes dispositivos
- **Consistência visual** que mantém elementos da interface que possuem a mesma função com o mesmo estilo nos diversos dispositivos
- **Consistência conceitual** que unifica a representação das diferentes tarefas em diferentes contextos
- e **Consistência de conteúdo**, sincronizando o conteúdo em diferentes dispositivos

Ainda não existe uma resposta fechada para qual tipo de consistência é mais apropriada para cada contexto. Os *designers* de interação criam bibliotecas de elementos e *guidelines* visuais para manter a consistência visual dentro de um projeto, mas no momento em que as telas deixam de ser as únicas interfaces de um projeto, outras formas de consistência devem ser buscadas. É preciso pensar em como criar consistência nos gestos, na linguagem e no visual, tornando a experiência multi-dispositivos contínua.

2.3.4 Interconexão de dispositivos

O penúltimo desafio levantado por Kuniavsky (2010) é o da interconexão de dispositivos. A comunicação entre os aparelhos é uma premissa fundamental da computação ubíqua, porém a complexidade tecnológica de fazer com que dois produtos digitais troquem informação um com o outro não é pequena.

Uma vez que os dispositivos conseguem se comunicar em um nível técnico, é necessário também que a experiência de uso faça sentido para essa relação. Algumas das formas levantadas por Kuniavsky (2010) para isso são:

- **Configuração manual.** O usuário deve configurar sozinho a relação dos dispositivos com manuais, telas de configuração e controladores de dispositivo (ou *drivers* de dispositivo). É um método que demanda tempo, documentação e conhecimento, além de ser mais limitado em termos de interconexões do que seria possível tecnicamente.
- **Aprisionamento tecnológico ou *vendor lock-in*.** O termo se refere à estratégia que faz com que os produtos de um mesmo fabricante sejam mais facilmente (ou exclusivamente) conectados entre si, desencorajando o uso de produtos de outro fabricante. Apesar de poder trazer benefícios para o negócio, é frágil e limitante do ponto de vista da interconexão de dispositivos, visto que dificilmente haverá uma maioria de usuários com todos os últimos lançamentos de uma única marca.
- ***Middleware*.** Os *middlewares* são *softwares* que traduzem padrões de comunicação digitais para conectar dispositivos que não seriam capazes de se conectar de outra forma. Sua limitação vem da necessidade dos desenvolvedores terem acesso aos protocolos dos dispositivos, e conseguirem traduzir esses protocolos de maneira efetiva. Além disso, é necessária uma atualização regular do *software* para que o mesmo continue funcionando.

- **Estratégia de introdução.** Os dispositivos podem comunicar sua existência passivamente, e os usuários então podem manualmente introduzir um dispositivo ao outro através de uma senha ou algum outro método de configuração. Esse é o caso, por exemplo, das conexões *bluetooth*, comuns atualmente.
- **Matchmaking.** Um dispositivo confiável introduz outros dois dispositivos um ao outro, provendo as informações necessárias para que uma comunicação confiável se estabeleça. Um exemplo desse tipo de comunicação são os cartões SIM (*subscriber identity module*, ou módulo de identificação do assinante) dos celulares, que fazem a conexão de um celular com uma rede de telefonia.

Para Kuniavsky, os problemas técnicos dessa interconexão de dispositivos será resolvida em um futuro próximo, e o que nos restará questionar é o papel das pessoas nesse ambiente conectado. O modo de ver mais cauteloso diz que os dispositivos de computação ubíqua irão precisar de intervenção humana constante para evitar o uso de recursos inesperados ou vazamento de informações pessoais. Criar um meio termo que equilibre segurança, utilidade e esforço do usuário final demandará reflexão sobre as relações e papéis que os sistemas de computação ubíqua desempenham nas vidas das pessoas. (Kuniavsky, 2010).

2.3.5 Discutir a complexidade e os limites do uso da tecnologia

Para Kuniavsky (2010), nesse novo campo que surge, identificar as limitações é tão importante quanto saber quando se manter dentro delas ou contestá-las. A definição desses limites tecnológicos e sociais é um desafio, pois a maioria das limitações não está na capacidade de uma dada tecnologia, mas na habilidade que o time de desenvolvimento tem para saber usá-la. Kuniavsky (2010) exemplifica dessa forma:

"Por exemplo, a maioria dos times provavelmente já passou pela situação onde um designer diz 'Eu gostaria de fazer X', e o engenheiro responde 'Isso será muito

difícil, se não impossível de fazer'. O designer então diz 'Bem, e se fizermos Y?', para o que o engenheiro responde 'Oh, isso é fácil'. Na perspectiva do designer, X e Y são muito similares, mas para o engenheiro, são problemas completamente diferentes." (KUNIAVSKY, 2010, p.29)

Isso significa que, apesar de alguns problemas realmente só poderem ser resolvidos com evoluções na tecnologia, existem também outros que o design pode contornar e que exigem apenas medidas triviais do ponto de vista da engenharia. Distinguir quais são os problemas de design e de engenharia é desafiador, mas existem medidas que permitem uma melhor compreensão das restrições tecnológicas, e que ajudam o time de desenvolvimento, como grupo, a lidar com elas. Alguns exemplos são: discussões constantes com os membros do time e partes interessadas, ter as documentações compartilhadas e gerar protótipos interativos, que permitam explorar como diferentes tecnologias podem trabalhar juntas. (Kuniavsky, 2010)

Kuniavsky (2010) pontua também que a tecnologia é sedutora e, apesar de acrescentar mais tecnologia nem sempre fazer um projeto melhor, é difícil definir quanto é o limite de "demais". Criar novas tecnologias é caro e é de interesse de um projeto apenas fazê-lo quando necessário. Conquanto, a necessidade de uma nova tecnologia pode vir de várias formas ao medir o valor de um produto: utilidade, usabilidade, custo-benefício, mas também atração emocional e apelo de marca. Isto posto, é importante reconhecer que algumas tecnologias não tem utilidade, usabilidade, custo-benefício e não agregam ao valor da marca a longo prazo, porém são atrativas para potenciais compradores. Assim sendo, saber quando respeitar as restrições de uma tecnologia é, em parte, entender que o objetivo de alguns produtos é puramente decorativo ou atrativo, e tomar decisões conscientes ao incluir esse tipo de tecnologia, sabendo quando faz sentido em longo prazo.

Todavia, os limites tecnológicos não são os únicos que devem ser considerados. Ao tratar de um mundo conectado, onde o lugar de uso de uma tecnologia pode ser qualquer, a possibilidade de uso compartilhado é uma premissa importante a ser ponderada. Celulares, por exemplo, dispositivos de uso

individual e pessoal, são usados em ambientes públicos, onde têm influência de outras pessoas.

Para Kuniavsky (2010), projetar tecnologias pensando em como elas serão compartilhadas entre as pessoas ajuda todos os tipos de designers, não apenas aqueles que estão projetando para um projeto que é explicitamente social.

O desafio desse atual paradigma, por tanto, passa por questões técnicas, mas também de adoção, mercado e experiência. A tecnologia pessoal deixa de habitar apenas o espaço privado e também vai para o espaço público, gerando mudanças no aspecto social do design e na responsabilidade do designer.

"E, é claro, escolhemos as coisas que colocamos ao nosso redor não apenas pelo que elas significam para nós, mas também pelo que elas significam para as outras pessoas. A maioria dos italianos possuem um telefone móvel, mas a maioria dos jovens Sicilianos não pode pagar pelas ligações. Eles ainda assim compram os telefones, porque ostentar um diz, de forma bastante explícita, 'Estou conectado com uma rede de familiares e amigos.' A função simbólica é tão importante quanto a função prática, talvez até mais." (Moggridge, 2007, p. 13, tradução nossa)

Ademais os 5 pontos tratados, uma porção significativa do desafio do design para a experiência de uso na computação ubíqua está em entender a prática do design, pois essa está mudando rapidamente e o papel dos criadores de novas tecnologias é profundamente transformador. Com isso, vem uma responsabilidade que precisa ser ativamente definida e gerenciada. (Kuniavsky, 2010)

2.4 As interfaces para a computação ubíqua

"No passado, aqueles que contruíam sistemas interativos propendiam a focar na tecnologia que fazia os sistemas possíveis, e não nas interfaces que permitiam o uso deles pelas pessoas. Porém, um sistema não está completo sem as pessoas que o utilizam. Goste ou não, pessoas - irrascíveis, exigentes, e frequentemente distraídas pessoas como nós - e seus objetivos são a razão de nossos sistemas, e nós devemos projetar para elas.

Criar para esse amplo novo espectro de humanidade é mais desafiador que inventar ferramentas especializadas para profissionais técnicos. Nossos usuários, justificadamente, não são preparados para gastar tempo aprendendo novos sistemas complicados. E ele não

são obrigados a usar nossos produtos: se eles não conseguem fazê-los funcionar, eles devolvem para a loja." (Moggridge, 2007, p. xii, tradução nossa)

Segundo Moggridge (2007), um sistema bem desenhado oferece *feedbacks* de confirmação, de forma que o usuário sabe o que ele fez no momento em que fez. Em um teclado, por exemplo, é possível saber o que está sendo escrito não apenas pelo *feedback* visual das letras que aparecem na tela conforme o usuário digita, mas também pelo retorno tátil das teclas quando são pressionadas. No caso retratado pelo autor, quando os teclados não eram silenciosos, era possível também identificar a tecla que era pressionada pelo retorno auditivo.

"Usando um processador de textos antigo para realizar uma tarefa repetitiva, eu costumava usar uma sequência de teclas que fazia "tetick, tick, tick-tick; tetick, tick, tick-tick". Se fizesse "tick, tetick, tock", eu sabia que havia cometido um erro". (Moggridge, 2007, p. 15, tradução nossa)

Por conta desses *feedbacks*, a digitação era mais rápida, pois era possível saber quando ocorreu um erro imediatamente, não sendo necessário contar apenas com a visão. Nesse exemplo, isso era viável por conta da interface física do teclado, mas novas soluções precisam ser pensadas quando migramos para outras formas de interação.

Para Moggridge (2007), além de *feedbacks*, outro recurso essencial para o desenho de uma boa interação é a navegabilidade: saber onde se está no sistema, o que pode ser feito ali, e para onde é possível navegar a partir desse ponto.

Uma forma de tangibilizar a navegação de uma interface é, por exemplo, com o uso de *breadcrumbs* e menus. O recurso *breadcrumbs*, uma das formas de mostrar a estrutura de organização de um site, é utilizado para sinalizar a sua localização em uma interface e o caminho realizado para chegar até aquele ponto. Outra forma de explicitar a navegação e a estrutura de um site é o menu, que mostra as opções de navegação.

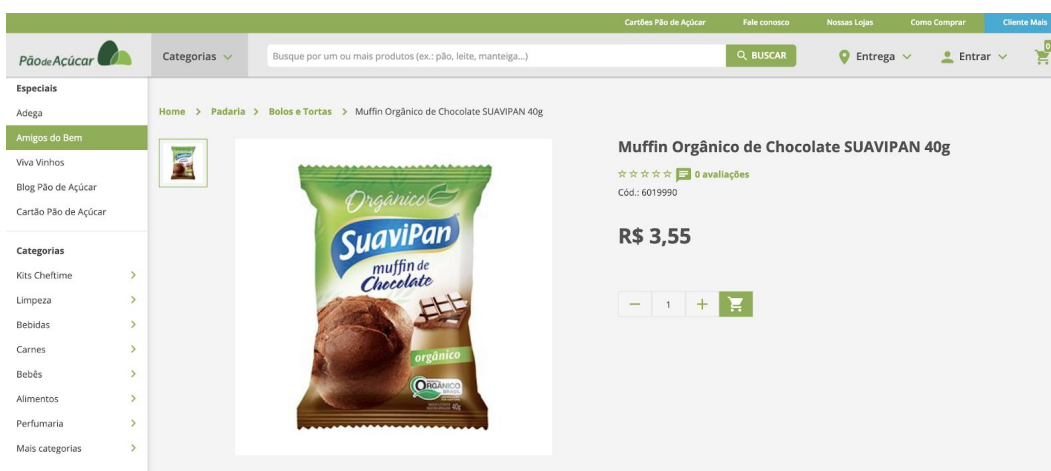


FIGURA 3 - Site do Supermercado Pão de Açúcar utilizando os recursos de *breadcrumbs* e menu

Moggridge(2007) levanta ainda um terceiro ponto de atenção ao desenhar interfaces: a consistência. Um tipo de interação ou comando deve funcionar da mesma maneira em todos os lugares dentro de um mesmo sistema. Isso permite que você não se perca facilmente na interface e que cometa menos erros.

Ao se desenhar um sistema ou dispositivo computacional, a tarefa do designer de interação não é apenas pensar apenas na aparência da interface, mas em como ela se comporta. A qualidade da interação entre o usuário e o sistema é, em parte, responsividade: se a resposta é adequada, se a performance é boa ou se é lenta, se a interação é clara. No entanto, é também entender se a forma de manipular um controle, além de dar um retorno para o usuário sobre o que ele está fazendo, é também satisfatória. (Moggridge, 2007)

Kuniavsky (2010) afirma que a velocidade na evolução da tecnologia e as mudanças nos padrões sociais fez com que identificar as melhores práticas do design da experiência de uso fosse difícil. Os produtos provenientes da computação ubíqua são híbridos de *hardware*, *software* e serviços.

Quando interagimos com objetos do dia-a-dia, nós não passamos muito tempo pensando na interação, mas focamos na tarefa que estamos realizando.

Interações intuitivas minimizam o esforço consciente de manipular um sistema, deixando o usuário livre para focar em seu objetivo. (Moggridge, 2007)

Weiser (1994) defende o computador do futuro como o computador invisível que executa sua tarefa sem tomar o foco. A pessoa que o utiliza deve ter o foco em seu propósito, não na ferramenta que ela usa para executar a tarefa. Para ele (Weiser, 1991), não é o excesso de informação sozinho que torna os computadores frustrantes e desperta emoções negativas. Para ele, a "invisibilidade" é um fator importante para tornar um produto agradável.

"Há mais informação disponível na ponta de nossos dedos durante uma caminhada no bosque que em qualquer sistema de computador, ainda assim as pessoas acham um passeio entre as árvores relaxante e computadores frustrantes. As máquinas que se encaixam no ambiente humano, ao invés das que forçam os humanos a se encaixarem no ambiente delas, irão fazer usar um computador tão revigorante quanto uma caminhada no bosque." (WEISER, 1991, tradução nossa)

Para exemplificar esse fato, Weiser (1994) usa o exemplo do óculos: ao usá-lo, a pessoa vê o mundo e não a ferramenta, o óculos em si. Ele serve o seu propósito e completa a tarefa sem tomar o foco, sem invadir a consciência.

"Os óculos são boas ferramentas - você vê o mundo, não os óculos. O homem cego que tateia o mundo com seu bastão sente a rua, não o bastão. É claro, as ferramentas não são invisíveis por si só, mas partes do contexto de uso. Com prática o suficiente, nós podemos fazer coisas aparentemente difíceis desaparecerem: meus dedos conhecem seis comandos de edição no computador que minha mente consciente há muito tempo esqueceu. Porém, boas ferramentas aumentam o caráter de sua invisibilidade." (WEISER, 1994, p.7-8, tradução nossa)

Tirar a interface do foco, no entanto, não significa não tratá-la com afinho. Ao contrário, ela deve ser muito bem pensada e planejada para que haja interesse por parte do usuário, porém não deve atrapalhá-lo na execução de suas atribuições. Norman (2008), defende que o atrativo desperta afetividade positiva, o que incentiva a pessoa a observar o todo, procurar novas possibilidades, explorar a criatividade. Ele traz a questão de que muitas vezes esse caráter emocional é o que faz um produto ser bem sucedido.

Greenfield (2006) dizia que quando os objetos do dia-a-dia tivessem a habilidade de reconhecer o ambiente e guardar informações que refletissem sua proveniência, localização, condição e histórico de uso, e quando esses pudessem compartilhar informações entre si, a relação com tais objetos seria modificada. Ele aponta que uma das divergências mais claras entre a computação ubíqua e os computadores pessoais é a demanda do primeiro por novas modalidades de *input* além dos tradicionais teclado, tela, *trackball*, *touchpad* e *mouse*, pois essas formas de interação não atendem às necessidades desse cenário. As diferentes funcionalidades distribuídas em objetos e no ambiente pedem por novas interfaces e maneiras para as pessoas comunicarem suas demandas e desejos para os sistemas computacionais em seu entorno.

Apesar de muitas vezes o futuro ter previsto um cenário de muitas telas e *displays* de informação na ficção, o caminho hoje é abordar novas formas de interação não-visuais. Dessa forma permitindo um ponto de equilíbrio entre atratividade, que gera conforto e afeto positivo, e invisibilidade enquanto ferramenta, sem tomar o foco da situação e permitindo ao usuário atingir seu propósito.

Dentre as novas formas de interação que estão sendo usada para tratar desse novo cenário, temos a realidade virtual, a realidade aumentada, a realidade mista, as interações gestuais no espaço físico e também a voz, dentre outras. Para Weiser (1994), a ideia de realidade virtual (também chamada VR, *virtual reality*) é posta à prova a medida que ela se torna o centro de toda a atenção do usuário.

O mesmo problema recai sobre a realidade aumentada e mista. Ao colocar a interface no mundo, corre-se o risco de tornar tudo uma interface, e de exagerar na quantidade. Quando a solução está em tornar tudo uma interface, foca-se novamente na ferramenta e não na atividade que se deseja realizar.

Essas interfaces devem ser estudadas e trabalhadas, pois abrem muitas possibilidades. No entanto, não são usadas largamente, pois ainda não são muitas as tarefas que se beneficiariam desse tipo de interação. Também, ainda não são

acessíveis em financeiramente e apresentam desafios tecnológicos que precisam ser superados - mas devem sê-lo em breve, com as evoluções e avanços dessa área.

As interfaces de voz, no entanto, tem aumentado seu alcance e se sustentam na fala, habilidade que os seres humanos desenvolvem desde novos. Por serem passíveis de serem acessados em celulares de baixo custo com acesso a internet, as interações por voz já são utilizadas no cotidiano das pessoas e se mostram importantes de serem investigadas para esse contexto da computação.

2.5 Resumo do capítulo

A evolução da computação permitiu que os dispositivos computacionais se espalhassem em objetos cotidianos e se desvinculassem da necessidade de telas e *displays* visuais de informação. Essa evolução se insere no contexto da computação ubíqua.

Um dos desafios desse cenário é a demanda por novas formas de interação além dos tradicionais teclado, tela, *trackball*, *touchpad* e *mouse*. Essas novas interfaces devem seguir os mesmos princípios convencionados para o *design* de uma interface visual: prover *feedback*, fácil navegabilidade, consistência e atratividade, sem perder sua função de ferramenta.

Além desse desafio, também é necessário superar outros obstáculos:

- **Introdução bem sucedida de novas tecnologias e experiências.** Como toda nova forma de interação compete com outra existente, é preciso que seus benefícios e diferenciais estejam claros para justificar o investimento no seu desenvolvimento e para propiciar sua adoção;
- **Projetar para a Adaptação.** Já que os produtos possuem uma vida maior que seu período de "popularidade", deve-se criar experiências que

enderecem os desejos e as necessidades mais imediatas das pessoas, mas que também permitem a sua adaptação no futuro;

- **Experiência de interação contínua em diferentes dispositivos.** Quando os dispositivos são vários e novas formas de interação surgem a todo momento, torna-se ainda mais importante saber como manter a consistência, seja ela operacional, visual, conceitual ou de conteúdo;
- **Interconexão de dispositivos.** Uma vez que os diferentes dispositivos computacionais conseguem se comunicar em nível técnico, premissa fundamental da computação ubíqua, é necessário questionar o papel das pessoas nesse ambiente conectado. Sabendo que a necessidade de intervenção humana constante para evitar o uso de recursos inesperados ou vazamento de informações pessoais se tornaria um fardo, é preciso chegar ao equilíbrio entre segurança, utilidade e esforço do usuário final;
- **Discutir a complexidade e os limites do uso da tecnologia.** Muitos problemas que aparentemente são de engenharia podem ser contornados com soluções de design, mas às vezes uma tecnologia aparentemente sem funcionalidade pode agregar valor emocional para uma marca. Visto que desenvolver novas tecnologias é caro, torna-se ainda mais importante entender o objetivo de um produto para saber quando incluir essa complexidade em um projeto;

3

Os Assistentes Virtuais Personificados e a fala

Conforme novas tecnologias e formas de interação surgem, criam-se também novos sistemas, como os assistentes virtuais. Eles podem ser chamados por muitos nomes, como agentes inteligentes (Weiser, 1994), assistentes pessoais virtuais (VPAs, *virtual personal assistants*), assistentes pessoais inteligentes, assistentes digitais pessoais, assistentes móveis ou assistentes de voz (MCTEAR; CALLEJAS; BARRES, 2016).

Os assistentes virtuais são sistemas que realizam tarefas e oferecem serviços variados. Atualmente, qualquer pessoa com conexão à *internet* e um celular recente, tanto *iOs* da Apple quanto *Android* do Google, pode acessá-los sem precisar recorrer a nenhum tipo de instalação. Eles podem auxiliar em uma gama enorme de funções do dia-a-dia, como buscas, organização de agenda, anotações, previsão do tempo, ajuste de despertadores e outros.

"Muitos desses VPAs (assistentes pessoais virtuais) ajudam os seus usuários desempenhando um diverso número de tarefas em seus smartphones, como obter informações usando pesquisa de voz, encontrar restaurantes locais, entregar direcionamentos de navegação, programar o despertador, atualizar a agenda, e engajar em conversas. Outros desempenham funções mais especializadas como monitorar a saúde, e preparar *drinks* e receitas." (MCTEAR; CALLEJAS; BARRES, 2016, p.11, tradução nossa)

Os assistentes virtuais têm a voz e o texto como principais métodos de entrada de dados (*input*). Para essa pesquisa, vamos considerar os seguintes dois tipos de interface:

- interfaces por **comando e controle**. Desencadeiam uma dada função a partir de *inputs* pré-configurados (limitados) por voz, texto, ou até DTMF (*dual-tone multi-frequency*, ou tons de duas frequências) no caso dos assistentes virtuais de *telemarketing* e *call center*. Geralmente, não há uma

ação de fala que dure mais de um turno, ou seja não há referência a enunciados anteriores do diálogo, pois cada fala é uma nova interação de comando. "Digite zero ou fale 'atendente' para ser transferido".

- e interfaces **conversacionais**. Usam inteligência artificial e aprendizado de máquina para permitir um indefinido número de *inputs* por fala ou texto e, a partir disso, interpretam uma intenção e desencadeiam uma dada função. Podem permitir conversas por mais de um turno, ou seja, resgatam o que foi dito anteriormente no mesmo diálogo. Esse é o caso dos assistentes virtuais encontrados nos celulares e *smartspeakers* atuais, como a Google Assistente, a Siri, a Cortana e a Alexa.

Weiser (1994) questiona a ideia de agente inteligente, levantando que um computador não deveria ser, necessariamente, semelhante a uma pessoa. Para ele, como os relacionamentos humanos não são simples e possuem seus próprios problemas, não deveríamos mirar na comunicação humana como modelo de interação ideal com um computador.

"Por que os computadores deveriam ser parecidos com um ser humano? Aviões são como pássaros, máquinas de escrever como canetas, alfabetos como bocas, carros como cavalos? As interações humanas são tão livres de problema, mal entendidos e ambiguidades que representam uma interface desejável? Indo mais além, construir e manter um time de pessoas que funcione tranquilamente, mesmo uma dupla de pessoas, toma muito tempo e atenção. Um computador que exige que eu fale com ele, dê comandos, ou mesmo tenha um relacionamento (ou pior, seja íntimo) é um computador que está muito no centro da atenção." (WEISER, 1994, p. 7-8, tradução nossa)

Apesar da provocação, Weiser (1994) acredita que os assistentes inteligentes e a voz como *input* são tecnologias importantes a serem estudadas. Ele apenas sugere que elas sirvam a uma gama limitada de atividades por serem interfaces no domínio das interações conscientes e, por tanto, não se deve esperar fazer tudo através da voz.

Em ambientes públicos, por exemplo, onde queremos falar e interagir com outras pessoas, um assistente pode não ser bem-vindo. Nesses casos, o computador pode ser um aliado para tomar notas, para procurar informações sobre

a conversa, para consultar e alterar agendas... No entanto, isso deve ser feito de maneira não obstrutiva, pois se fosse necessário consultar um computador através da voz nesses momentos, ele se tornaria o centro da interação. (Weiser,1994)

No entanto, os assistentes não usam apenas a interface de voz como forma de interação; alguns até não a utilizam, como é o caso da maioria dos *chatbots*, cuja interface simula a troca de mensagens de texto. Além da voz e do texto, quando a disposição de uma tela ou outra interface visual, os assistentes costumam manipular esse recurso para aprimorar suas respostas. Na Google Assistente, por exemplo, ao perguntar pelo tempo, o usuário recebe uma imagem junto com o retorno auditivo, mostrando a previsão hora a hora (figura 4).



FIGURA 4 - Google Assistente exibindo informações visuais complementares

Em outros casos, como o do *StatMuse* (figura 5), as imagens são utilizadas para dar vida ao personagem criado para humanizar o assistente.

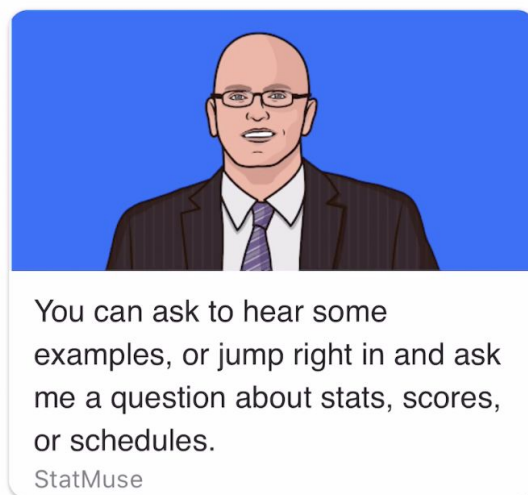


FIGURA 5 - Uso de imagem para personificar um assistente

Mesmo que os assistentes virtuais sejam cada vez mais comuns, os estudos acerca do uso, opinião e adoção dos mesmos ainda são limitados, muitas vezes tendo foco na tecnologia. As razões que motivam a experimentação de uma nova tecnologia são várias, mas os motivos que fazem que um usuário a aceite e adote no seu dia-a-dia vão além das suas capacidades técnicas e entram em uma dimensão de utilidade e emoção.

"Mesmo que um produto seja avançado em questões de tecnologia, ele não irá ter sucesso caso não seja aceito e adotado pelos usuários. Até recentemente, parecia que os usuários haviam parado de usar seus VPAs após um período inicial de experimentação. Em alguns casos, eles haviam encontrado problemas e erros no reconhecimento da fala, e então voltaram a usar outros *inputs* com os quais estavam mais acostumados e com os quais tinham maior acurácia. Em outros casos, os usuários se encantavam em dizer coisas bobas para os seus VPAs e ver que tipo de resposta teriam." (MCTEAR; CALLEJAS; BARRES, 2016, p.18, tradução nossa)

Os Assistentes Virtuais Personificados, que para esse trabalho trataremos como AVPs, são um tipo de agentes inteligentes. Em termos simples, eles desempenham tarefas que envolvem diferentes variáveis, e às vezes aplicações, como que secretários de seus usuários. Eles usam interfaces conversacionais por voz e incorporam uma persona (um personagem com características e personalidade próprias), de forma que a linguagem deles se torna mais natural e humanizada. Esses assistentes podem ser encontrados em diversos sistemas como,

por exemplo, embarcados nos principais OSs (Sistemas operacionais) de *smartphones* atuais.

São exemplos de AVPs a Siri da *Apple*, a Google Assistente do *Google*, a Cortana do *Windows* e a Bixby nos celulares da *Samsung*. Para essa pesquisa, daremos foco à Siri e à Google Assistente, por se tratarem de AVPs de mais fácil aquisição.

3.1 A voz e a fala como formas de interação

Segundo Gomes (2007), a comunicação "não é um aspecto acidental da vida humana, mas sim algo essencial e natural". Ela é o processo no qual ocorre a troca de informações, e o compartilhamento de experiências ou de sentimentos, por meio de uma mensagem que é passada de um emissor para um receptor através de um sistema de signos (palavras, gestos, sons, imagens, etc).

Quando as pessoas se engajam em uma conversa, elas transmitem mais que o significado das palavras emitidas. O discurso também passa o estado emocional e os aspectos da personalidade do locutor. A interação cara a cara permite ainda a transmissão de significado nos comportamentos não-verbais, como expressão facial, gestos e postura corporal. (MCTEAR; CALLEJAS; BARRES, 2016)

Ademais, a fala é o meio de comunicação humano fundamental. Mesmo quando a escrita, as expressões faciais, a linguagem de sinais e outras formas de comunicação são igualmente expressivas; as pessoas de todas as culturas (que conseguem escutar) persuadem, informam e constroem relacionamentos primariamente pela fala. (NASS; BRAVE, 2007)

Para Grice (1975), o que foi comunicado não está, necessariamente, contido no discurso da conversa, pois existem mensagens que podem ser subentendidas dado um contexto. Isso se dá porque, mesmo tendo conhecimento da língua falada

pelo locutor, um ouvinte precisa conhecer as circunstâncias de uma frase para ser capaz de interpretar de forma adequada o enunciado.

Segundo ele (GRICE, 1975), uma conversa não é composta por uma série de constatações sem relação entre si. Até certo ponto elas partem de um esforço cooperativo, onde cada participante reconhece um propósito comum ou, pelo menos, onde ambos aceitam mutuamente uma direção. Esse objetivo ou direcionamento pode ter sido estabelecido desde o início da conversa, ou pode ter evoluído durante a mesma; pode ser um objetivo determinado, ou um indefinido como durante uma conversa informal entre amigos. Em todos os casos, algumas atitudes não são adequadas para uma conversa, dado que o esperado é que cada pessoa faça sua contribuição no momento certo, e de acordo com o propósito ou na direção da conversa que se está envolvido. Grice (1975) chama esse fundamento de “**Princípio Cooperativo**”.

Para que uma conversa seja bem sucedida dentro do princípio cooperativo, e para que haja uma comunicação efetiva, Grice (1975) definiu quatro axiomas:

- **Quantidade** se refere ao montante de dados que é provido. Um discurso deve trazer tanta informação quanto é necessária para o propósito de uma troca, e não menos. Existe debate quanto a trazer mais que o essencial: pode ser defendido que é apenas uma perda de tempo; porém o excesso de informação pode levar à confusão do ouvinte, ao achar que há uma razão para o que foi dito demais.
- **Qualidade** trata da veracidade do que foi falado. Presa-se pela confiança: não se deve expressar algo que não se acredita ser verdade, ou que não se tem evidências adequadas.
- **Conexão** diz respeito a relevância do que se fala para a conversa. As informações ditas devem ser de interesse para o momento da conversa. No entanto, existem diferentes tipos e focos de relevância; a conversa permite uma mudança de direcionamento durante seu progresso.

- **Forma** aborda a maneira como algo é expressado. Deve haver clareza nos enunciados, não ser prolixo, não falar ambiguidades, evitar termos desconhecidos e ter uma fala ordenada (em uma disposição compreensível).

Apesar desses princípios serem básicos, eles podem ser deixados de lado quando entram em conflito um com o outro. Por exemplo, uma pessoa que não sabe detalhar uma informação pode optar por contar para a outra menos que o necessário (quantidade), visando não passar dados errados (qualidade). Grice (1975) demonstra isso com o seguinte exemplo:

"A está planejando junto com B o itinerário de uma viagem para a França. Ambos sabem que A quer ver seu amigo C, caso isso não prolongue muito sua viagem:

A: Onde C mora?

B: Em algum lugar no sul da França.

(Comentário: Não há razão para pensar que B está desconversando, mas sua resposta é, como ele bem sabe, menos informativa que o necessário para satisfazer as necessidades de A.

Essa violação do primeiro axioma de Qualidade pode ser explicada apenas pela suposição de que B está ciente que, para ser mais informativo, ele precisaria dizer algo que infringe o axioma de Qualidade 'Não diga algo que lhe falta evidências adequadas'. Então, B deixa subentendido que ele não sabe em que cidade C vive.)" (GRICE, 1975, p.51-52, tradução nossa)

Existem também casos onde os axiomas são propositalmente desconsiderados para trazerem implicações. Ao deixar de lado Quantidade e informar menos que o esperado, por exemplo, pode-se pressupor que não há mais a se falar sobre um assunto. É o que Grice (1975) demonstra no seguinte exemplo:

"A está escrevendo um depoimento sobre seu aluno que está se candidatando para um trabalho em filosofia, e sua carta prossegue assim:

'Prezado senhor, o domínio da língua inglesa de Sr. X é excelente, e sua participação nas aulas é regular. Atenciosamente, etc.

(Comentário: A não está fugindo da conversa, visto que, quisesse ele não cooperar, não escreveria o depoimento. Ele também não pode ser incapaz de dizer mais por ignorância, pois Sr. X é seu aluno; ademais, ele sabe que é desejado mais informação que o que ele deu.

Ele deve, portanto, querer transmitir informações que ele reluta em escrever. Essa teoria é sustentável apenas dada a suposição de que ele acredita que o Sr. X não é

bom em filosofia. Isso, então, é o que ele está implicando.)" (GRICE, 1975, p.52, tradução nossa)

Pode-se também descumprir o princípio de Qualidade ao se usar de ironia, metáforas e hipérboles. O axioma Conexão é mais raro de ser violado, porém é possível, como Grice (1975) exemplifica:

"Em uma refinada festa de chá, A diz que a Sra. X é uma mal-humorada. Há um momento de silêncio, e então B diz 'O tempo tem estado bastante agradável neste verão, não é?'. B se recusou a dizer algo relevante para a observação anterior de A. Dessa forma, ele implica que a observação de A não deveria ser discutida e, talvez mais especificamente, que A cometeu uma gafe social." (GRICE, 1975, p.52, tradução nossa)

Outrossim, o princípio da Forma pode ser deixado de lado ao querer passar uma ambiguidade, ou ao querer ser hermético para que uma criança no mesmo ambiente não consiga entender a conversa de um adulto. Todos os axiomas podem ser desconsiderados em alguma situação específica, geralmente dada a intenção de deixar algo subentendido.

Da mesma forma, Nass e Brave (2007) demonstram que a escolha dos termos de uma frase é capaz de carregar muitas informações. Por exemplo, a forma que uma pessoa comunica a não compreensão de algo pode passar diferentes significados dependendo das palavras escolhidas:

- **"Desculpe, eu não entendi."** toma responsabilidade do não entendimento para si;
- **"Seja mais articulado."** passa a culpa para a outra pessoa;
- E **"A torre de transmissão está com problemas. Você poderia repetir?"** usa um terceiro elemento como escapatória.

A comunicação através da fala é complexa, mas o ser humano evoluiu para entendê-la, e para extrair aspectos sociais e informações subentendidas. Porém, atualmente, emergiram tecnologias capazes de produzir e compreender vozes. As VUIs (*Voice User Interfaces*, ou interfaces de voz) "complementam, e por vezes

substituem, as interfaces gráficas, libertando os usuários das limitações das WIMPs (*windows* ou janelas, ícones, menus e ponteiros)". (NASS e BRAVE, 2007)

Na computação ubíqua, onde há acesso a informação em qualquer lugar e a todo momento, a fala serve como forma de interação para aqueles cujos olhos ou mãos estão ocupados em outras tarefas, ou para aqueles que não podem ler ou digitar, como crianças, cegos e incapazes. Todavia, essa interface cria uma questão interessante: o entendimento que se teve até então, de que as vozes fazem parte intrinsecamente do universo da sociabilização, precisa ser redefinido. Isso porque as pessoas agora conversam não apenas umas com as outras, mas também com a tecnologia. (NASS; BRAVE, 2007)

Nass e Brave (2007) mostram que o cérebro humano não costuma fazer distinção entre falar com uma pessoa e falar com uma máquina, mesmo quando a produção e a compreensão de fala da mesma é de baixa qualidade. As pessoas usam as mesmas partes do cérebro ao interagir com computadores e humanos.

"As pessoas tiram conclusões a respeito das vozes geradas por computador e adequam seu comportamento para com elas usando as mesmas regras e os mesmos atalhos que os utilizados para interagir com outras pessoas. Essas tecnologias, da mesma forma que a fala com outras pessoas, ativam *todas* as partes do cérebro associadas com a interação social." (NASS; BRAVE, 2007, n.p, tradução nossa)

Como consequência, ao projetar interfaces de voz, os *designers* de interação devem levar em consideração que estarão criando para pessoas acostumadas com "extrair tanta informação social quanto possível da fala, e habituadas em usar essa informação para guiar suas atitudes e comportamentos" (NASS; BRAVE, 2007, n.p, tradução nossa).

De acordo com Nass e Brave (2007), as pessoas são capazes de interagir com falantes de outras línguas, crianças pequenas, conversar através de uma linha telefônica com ruído, e escutar áudio de baixa qualidade em caixas de som ruins.

Essas situações possuem as mesmas limitações, ou até mais, que as encontradas nas tecnologias de hoje.

Apesar dos benefícios da voz como interface, ela pode não ser a mais adequada para todos os sistemas, pois ainda está sujeita a erros de reconhecimento que podem gerar frustração para o usuário. Além disso, existem tarefas para as quais o uso da voz não é aconselhável, como apresentar listas de dados ou muitas opções. Uma das soluções, como mostra Gomes (2007), é passar a combinar a voz com outras modalidades como gráficos e textos, criando interfaces multimodais. A tendência de diminuição do tamanho das telas nos dispositivos "torna igualmente atrativa a possibilidade de interação multimodal, pois as dimensões reduzidas dos aparelhos colocam dificuldades de usabilidade". (GOMES, 2007, p.2, tradução nossa)

3.2 Interfaces conversacionais

Segundo McTear, Callejas e Barres (2016), o termo conversa costuma se referir a uma interação informal onde há a transmissão de informações e opiniões por meio da fala, e onde o objetivo primário é a manutenção das relações sociais. Ela se contrasta com o discurso por tratar de uma interação onde há, necessariamente, uma troca.

Uma interface conversacional permite uma interação de conversa entre humanos e computadores, ou seja, um usuário consegue interagir com aplicações computacionais usando a voz de maneira mais ou menos natural. O termo "interface conversacional" também é usado em oposição a outros sistemas que usam a fala como interação, porém sem o suporte de uma comunicação mais humana e espontânea. Nesse caso, a interface conversacional é aquela que não usa comandos de voz através de frases curtas e palavras de ação, mas sim propõem uma interação mais próxima a um diálogo, onde há troca de informações. (MCTEAR; CALLEJAS; BARRES, 2016)

Nas interfaces de comando e controle (*command and control*) ou nos sistemas de ditado por voz, não existe um diálogo propriamente dito, "o sistema limita-se a executar os comandos ou a transcrever o que o utilizador disse". (GOMES, 2007, p.2). Dessa forma, as interfaces conversacionais podem ser consideradas aplicações avançadas das tecnologias de fala, incluindo áreas como o reconhecimento e síntese de voz, processamento da linguagem e gestão de diálogo.

Para Cathy Pearl (2016), a definição de interface conversacional é mais rígida. Já são consideradas interfaces de comando e controle aquelas que o usuário precisa indicar quando vai iniciar uma interação, seja a invocação por meio de uma fala como "E aí, Siri?" ou por um botão. Para essa pesquisa, esse não será tratado como um fator determinante, no lugar, usaremos as definições como relacionadas acima.

As interfaces conversacionais fazem parte do repertório da ficção científica desde há muito tempo, mas tardou para que saíssem da ficção (e do papel) e estivessem disponíveis para o público. Segundo McTear, Callejas e Barres (2016), em 1987, a *Apple* lançou um vídeo conceito de um dispositivo chamado *Knowledge Navigator*, mas apenas em 2011 as interfaces conversacionais deixaram de ser uma visão de futuro. É importante pontuar que as interfaces de voz já existiam nesse contexto, porém não eram interfaces conversacionais.

"Porém foi apenas em 2011 que essas visões foram concretizadas com o lançamento da Siri, reconhecida como a primeira VPA ativada por voz.

A Siri e sistemas conversacionais similares só foram possíveis como resultado dos avanços da tecnologia e da crescente aceitação e adoção por parte dos usuários, como explicado em mais detalhes nas sessões seguintes." (MCTEAR; CALLEJAS; BARRES, 2016, p.16, tradução nossa)

Os autores retratam que houve um aumento recente no uso dessas interfaces. Eles defendem que esse crescimento se dá por fatores como os avanços na área de inteligência artificial, tecnologias da linguagem, evolução dos dispositivos, o

crescimento da *web* semântica e as melhorias na conectividade e nas redes sem fio. (MCTEAR; CALLEJAS; BARRES, 2016)

A inteligência artificial proporciona não apenas a melhora nas tecnologias de linguagem, onde se encontra o reconhecimento da voz e a fala, como também permite sistemas capazes de aprenderem sozinhos. A *web* semântica estrutura o conteúdo da *internet* e o faz ser capaz de ser lido por computadores, facilitando a compreensão do significado por trás do que está escrito, não se restringindo à leitura de forma literal e através de palavras-chave. (MCTEAR; CALLEJAS; BARRES, 2016)

A habilidade que os dispositivos têm para entender o que o usuário diz, e então realizar um ação, vêm da combinação de duas importante tecnologias: ASR (*automated speech recognition* ou reconhecimento automático de fala) e NLU (*natural language understanding* ou compreensão de linguagem natural). Com apenas o reconhecimento da fala, o sistema seria incapaz de realizar alguma ação com o que foi dito.

"Se alguém falasse com você em uma língua que você não conhece, você provavelmente conseguiria escrever foneticamente o que foi dito. Essa é a parte do ASR. Porém, você não faria ideia do que esse alguém quis dizer." (PEARL, 2016, p.13, tradução nossa)

De acordo com McTear, Callejas e Barres (2016), uma interface conversacional costuma funcionar da seguinte maneira ao receber um *input* de fala do usuário:

- O sistema reconhece as palavras que foram ditas (ASR, *Automatic Speech Recognition*, ou reconhecimento de fala)
- Interpreta as palavras, isso é, descobre o que o usuários queriam dizer e o que intencionavam ao falar os termos usados (SLU, *Spoken Language Understanding*, ou entendimento de linguagem falada)
- Formulam uma resposta ou, caso a mensagem recebida esteja intrincada ou incompleta, interagem com o usuário para buscar esclarecimento e pedir que

comuniquem as informações requeridas (*Dialog Management*, ou Gestão de diálogo)

- Controem uma resposta, que pode ser na forma de palavras, ou acompanhadas por informações visuais e outras em aplicações multimodais (*Response Generation*, ou geração de resposta)
- Falam e exibem a resposta (TTS, *Text-to-speech synthesis*, ou síntese de texto para fala)

Esse fluxo pode ser observado na figura 6.⁵

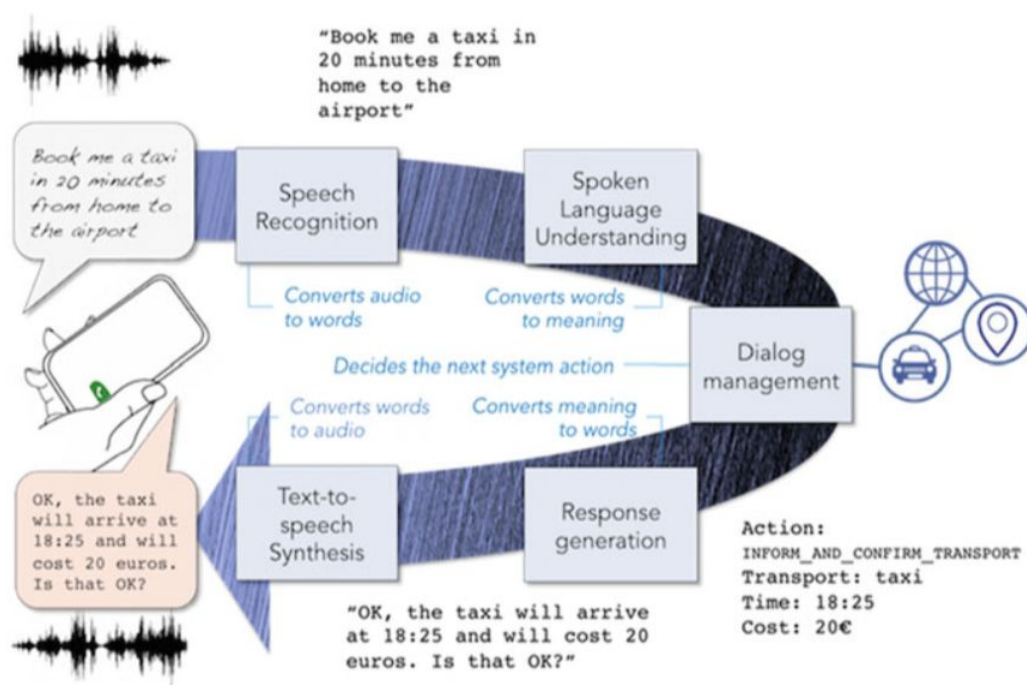


FIGURA 6 - Tecnologias envolvidas na interface conversacional (MCTEAR; CALLEJAS; BARRES, 2016, p.21)

Além de todas as evoluções tecnológicas, é de grande importância o interesse de grandes empresas, pois são elas que investem e tornam essas tecnologias acessíveis a grandes massas. Por razões de exemplo, é válido destacar

⁵ Existem diversas formas para que os sistemas realizarem o fluxo. Entrar em aspectos mais técnicos sobre o funcionamento de cada uma dessas etapas pede por um aprofundamento que não cabe no escopo da pesquisa de design.

atualmente, o investimento em interfaces conversacionais feito por empresas globais, como *Google, Apple, Microsoft, Samsung e Facebook*.

"Enquanto anteriormente o interesse em interfaces conversacionais para VPAs (Assistentes Pessoais Virtuais) era limitado para um pequeno nicho de companhias e entusiastas evangelizadores do sonho da inteligência artificial, agora muitas das maiores companhias no mundo estão competindo para criar seus próprios VPAs, como por exemplo, a Siri da Apple, o Google Now da Google, a Alexa da Amazon, a Cortana da Microsoft, o M do Facebook e o Duer da Baidu." (MCTEAR; CALLEJAS; BARRES, 2016, p.18, tradução nossa)

Porém, apenas ter a tecnologia que faz possível a interface conversacional não é o suficiente. Pearl (2016) levanta que ter reconhecimento de voz com alta acurácia resolve apenas uma parte do problema. Afinal, o que pode ser feito com essa informação? É preciso não apenas reconhecer as palavras, mas fazer algo que as pessoas realmente queiram. Para isso, é fundamental entender como as pessoas se relacionam com essa nova tecnologia.

"E no domínio dos sentimentos, é tão razoável se afeiçoar e amar coisas que são feias quanto o é não gostar de coisas que seriam chamadas de atraentes. As emoções refletem nossas experiências pessoais, associações e lembranças." (NORMAN, 2008, p.68)

3.3 Resumo do capítulo

A comunicação é o processo no qual ocorre a troca de informações e o compartilhamento de experiências ou de sentimentos, por meio de uma mensagem que é transmitida de emissor para um receptor através de um sistema de signos.

Qualquer conversa parte de um esforço cooperativo onde cada participante reconhece um propósito comum ou onde ambos aceitam mutuamente uma direção. Esse fundamento é chamado de "Princípio Cooperativo" e rege quatro axiomas de uma comunicação bem sucedida: Quantidade, Qualidade, Conexão e Forma. Apesar desses princípios serem básicos, eles podem ser deixados de lado quando entram em conflito um com o outro. Existem também casos onde os

axiomas são propositalmente desconsiderados para trazerem implicações, ou quando se faz uso de ironia, hipérboles e metáforas.

A comunicação é complexa, mas o ser humano evoluiu para entender e extrair mensagens, aspectos sociais e informações subentendidas da fala. Apesar disso, o cérebro humano não faz distinção entre falar com uma pessoa e falar com uma máquina.

No entanto, emergiram tecnologias capazes de produzir e compreender vozes, o que gera uma questão interessante: o entendimento que se teve até então, de que as vozes fazem parte do universo da sociabilização humana, precisa ser redefinido. Isso porque as pessoas agora conversam não apenas umas com as outras, mas também com a tecnologia.

Dentre as formas de interação por meio da fala, está a interface conversacional, que permite uma interação de conversa mais natural e espontânea entre humanos e computadores, e se opõem a outros sistemas como as interfaces de comando e controle. As interfaces conversacionais, por tanto, podem ser consideradas aplicações avançadas das tecnologias de fala, incluindo áreas como o reconhecimento e síntese de voz, processamento da linguagem e gestão de diálogo.

Apesar dos benefícios do uso da voz como interface, ela não é a mais adequada para todas as situações, pois além de estar sujeita aos erros de reconhecimento que podem gerar frustração para o usuário, existem tarefas para as quais o uso da voz não é aconselhável. Como alternativa para essas situações, uma das soluções é a de interfaces multimodais, onde se combina a voz com outras modalidades como gráficos e textos.

4

O residente do Rio de Janeiro e os Assistentes Virtuais Personificados que usam a voz como interface

Durante a pesquisa de revisão bibliográfica acerca dos temas de computação ubíqua, interação por voz, interfaces conversacionais e assistentes virtuais, muito foi levantado para definição do contexto. Essa revisão foi usada como referência para a pesquisa de campo com o público - residente do Rio de Janeiro -, de forma a apontar as particularidades do mesmo.

A pesquisa de campo foi realizada em duas etapas. A primeira consistia em um questionário online de perguntas fechadas, cujo intuito era entender a adoção e o uso de assistentes no Rio, bem como recrutar participantes para o grupo de foco.

A segunda etapa foi o grupo de foco, que visava levantar a aceitação e as opiniões do residente do Rio de Janeiro sobre os Assistentes Virtuais Personificados que usam a voz como interface. Além disso, nessa etapa foram mapeadas suas dificuldades e preocupações com a tecnologia, bem como seus pontos positivos e possibilidades para o futuro.

4.1 Questionário

O questionário é um método de investigação objetivo constituído por uma série ordenada de perguntas, que devem ser assinaladas individualmente e sem a presença de um entrevistador. O questionário pode ser feito a distância, sendo vantajoso por conseguir atingir um número grande de pessoas em pouco tempo.

Essa técnica permite o uso de perguntas fechadas, quando as opções de resposta são pré-definidas, ou de perguntas abertas, permitindo que o respondente redija sua própria resposta. Enquanto as perguntas fechadas possibilitam respostas

mensuráveis e precisas, as perguntas abertas permitem entender melhor o processo de pensamento do respondente, que pode qualificar e esclarecer respostas.

O objetivo da primeira etapa era entender de forma quantitativa a adoção e a experiência de uso dos assistentes, além de recrutar participantes para o grupo de foco. Era preciso coletar, por tanto, respostas claras e objetivas em quantidade, de maneira que o formato escolhido foi o de perguntas fechadas. As únicas respostas que deveriam ser redigidas eram as de solicitação de informações de contato para o recrutamento do grupo de foco (*e-mail* e telefone).

O perfil selecionado foi de moradores do Rio de Janeiro (estado) maiores de 18 anos. O filtro de local foi necessário, pois havia a intenção de recrutar esses respondentes para participar de um grupo de foco presencial que seria realizado no RJ. A maioria foi importante, pois menores de idade precisariam de aprovação de seu responsável legal.

Na construção do questionário, as primeiras perguntas foram filtros para restringir a pesquisa para o perfil e identificadores demográficos (local, idade, gênero e escolaridade). Quem se encaixava dentro do perfil selecionado para a pesquisa, respondia sobre o seu uso de internet e se havia utilizado algum assistente virtual nos três meses anteriores, de forma a levar para os próximos itens apenas quem conhecia o sistema e teria lembrança da experiência para opinar sobre o mesmo. Quem não havia tido contato recente já seria considerado nessa etapa como uma não adoção.

Os respondentes que haviam tido contato recente seguiam para as questões referentes ao uso, à opinião e às percepções gerais acerca do assistente virtual que haviam usado. O questionário mapeava qual o assistente utilizado, o dispositivo de acesso ao assistente, a finalidade de uso, a avaliação da experiência e a adoção do assistente. Após todas as perguntas, a pessoa deveria informar se teria interesse em participar do grupo e foco e, em caso afirmativo, eram solicitados seus dados de contato.

O questionário ficou disponível *on-line*, e foi divulgado para usuários comuns em mídias sociais, bem como presencialmente e através de cartazes no ambiente universitário. Ele foi distribuído entre abril e setembro de 2018 em grupos e páginas do *Facebook*⁶ de diferentes cursos e universidades do Rio de Janeiro, bem como grupos voltados para pesquisa, como o grupo **Classificados de pesquisa**. Além disso, o questionário foi divulgado com colegas da área de tecnologia (Globo.com e INFNET).

O questionário foi respondido por 200 pessoas, sendo 161 moradoras do Rio de Janeiro maiores de idade, dentre as quais 102 haviam utilizado algum assistente virtual que mimetizava a interação com uma pessoa nos três meses anteriores.

4.1.1 Características gerais dos respondentes

A maioria dos respondentes estava dentro do perfil de morador do Rio de Janeiro (166 pessoas), conforme gráfico 1.

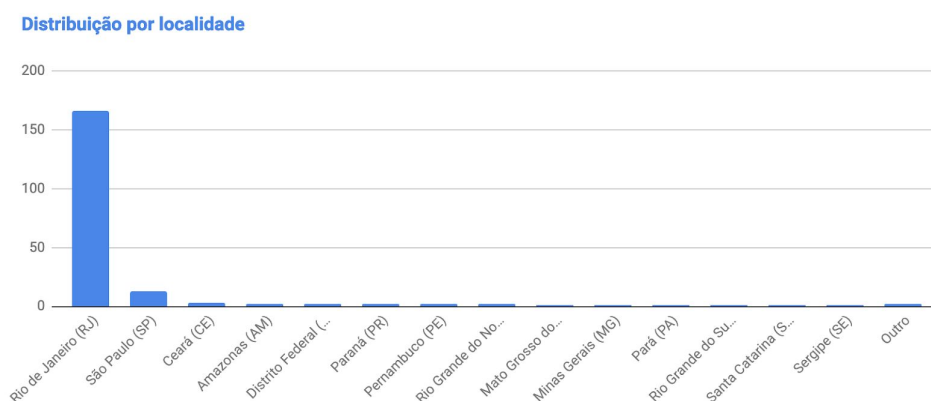


GRÁFICO 1 - Distribuição de respondentes por local de moradia

⁶ Na página **LEUI Laboratório de Ergodesign e Usabilidade de Interfaces PUC-Rio** do laboratório do curso de design da PUC-Rio, no grupo **Comunicação Visual Design - UFRJ** do curso de design da UFRJ, no grupo **Alunos e ex alunos Profa Alessandra Baiocchi** do curso de marketing da PUC-Rio, no grupo **ADM PUC-RIO** do curso de administração da PUC-Rio, no grupo **Dúvidas Design PUC-Rio** do curso de design da PUC-Rio, no grupo **PUC-Rio_Design** do curso de design da PUC-Rio, no grupo **Alunos do LEUI** do laboratório do curso de design da PUC-Rio, no grupo **Boulevard ESDI** do curso de design da UERJ, no grupo **Engenharia UFF** do curso de engenharia da UFF, no grupo **TI CEFET** do curso de tecnologia da informação do CEFET.

A média de idade dos respondentes foi de 30,6 anos com o mínimo de 16 e máximo de 76 anos. No Rio de Janeiro, a média é menor, 29,01 anos, também com o mínimo de 16 e máximo de 76 anos. A distribuição segue como demonstrada no gráfico 2.

Distribuição etária dos respondentes

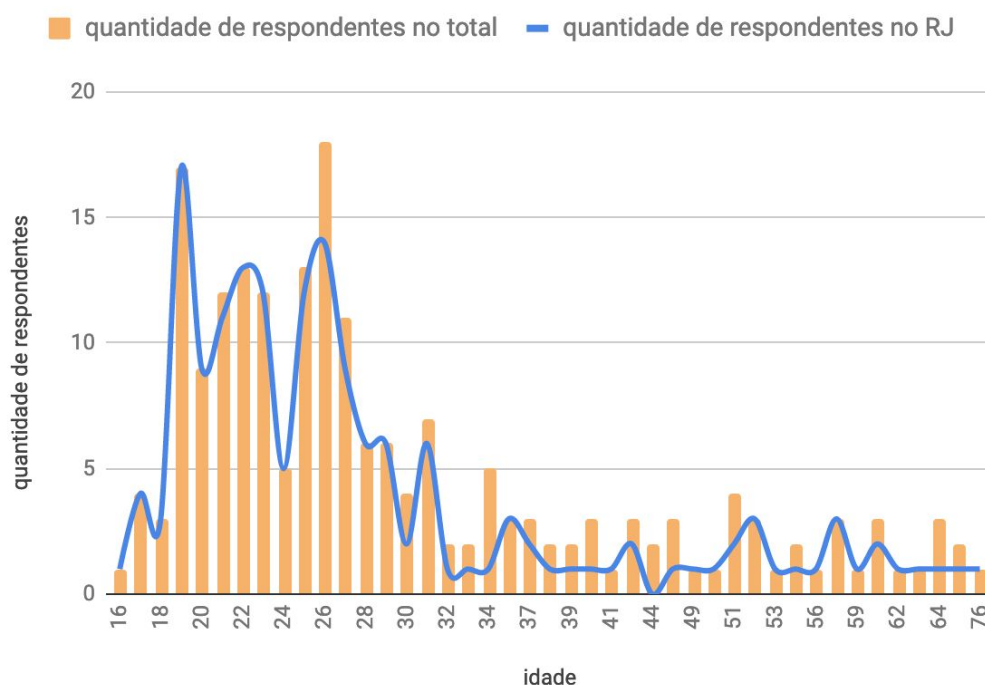


GRÁFICO 2 - Distribuição etária dos respondentes

A escolaridade dos respondentes manteve a mesma distribuição no geral e no Rio de Janeiro. O grau de instrução mínimo foi de ensino médio (ou segundo grau) completo, e o máximo foi pós-graduação ou especialização completo, como observado no gráfico 3.

Escolaridade dos respondentes

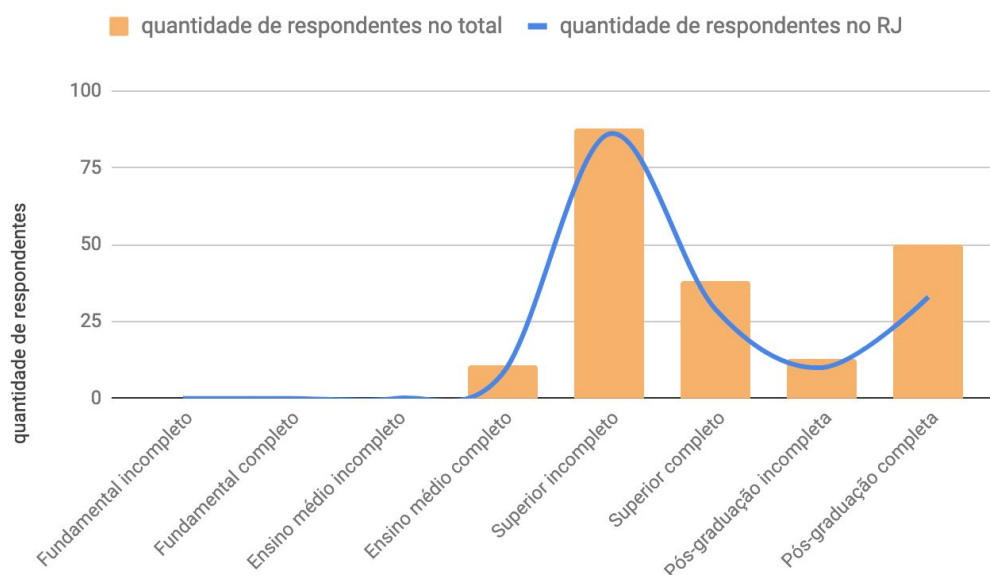


GRÁFICO 3 - Escolaridade dos respondentes

A distribuição por gênero foi equivalente. No total, foram 49,5% de homens e 48,0% de mulheres, e no Rio de Janeiro foram 49,4% de homens e 47,5% de mulheres. O gráfico 4 demonstra essa distribuição. 5 respondentes optaram por não declarar seu gênero.

Gênero dos respondentes

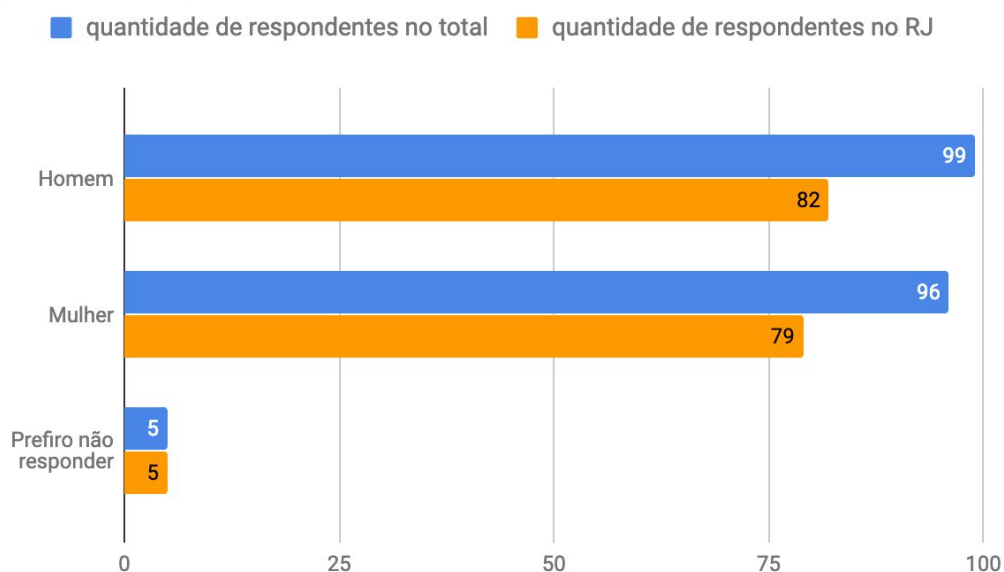


GRÁFICO 4 - Distribuição por gênero

4.1.2 Características do perfil dos respondentes

O perfil alvo da pesquisa foi de moradores do Rio de Janeiro com mais de 18 anos. Esse grupo de respondentes foi representado pelas características descritas a seguir.

A idade média dos respondentes que correspondiam ao perfil era de 29,4 anos, com a maioria deles na faixa entre 19 e 31 anos, como no gráfico 5.

Distribuição etária do perfil

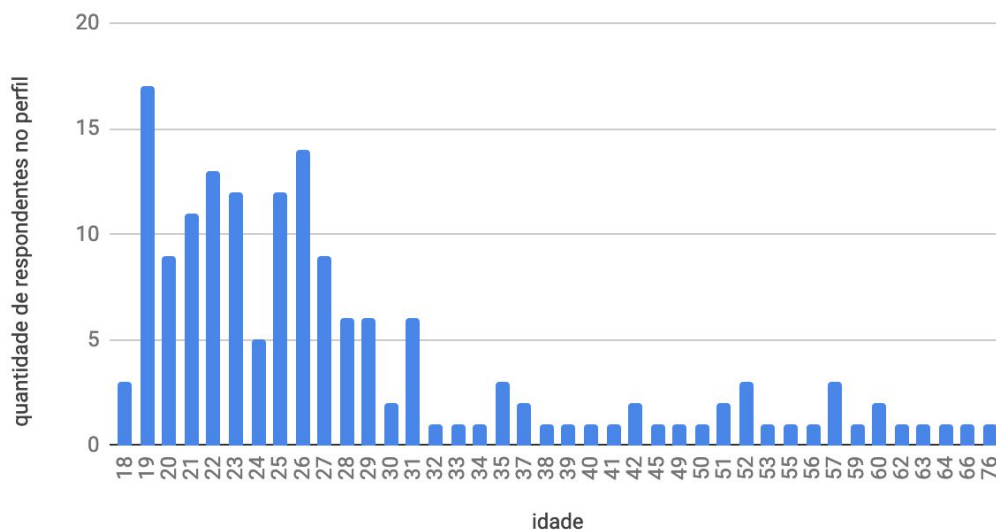


GRÁFICO 5 - Distribuição etária entre respondentes do Rio de Janeiro maiores de idade

O grau de escolaridade era, principalmente, ensino superior incompleto, como mostra o gráfico 6.

Escolaridade do perfil

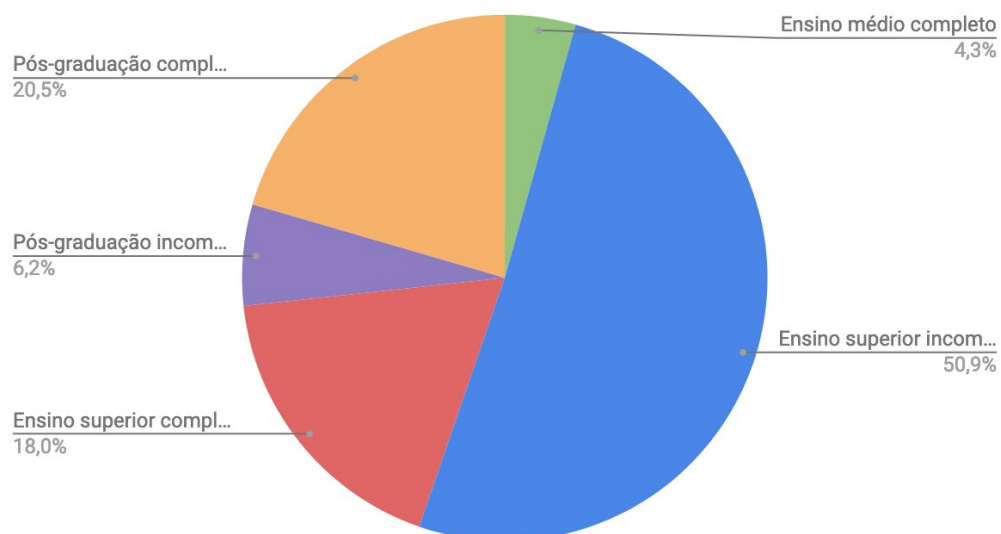


GRÁFICO 6 - Escolaridade dos respondentes maiores de idade do Rio de Janeiro

Os respondentes do perfil se mostraram bem distribuídos quanto à declaração de gênero, conforme o gráfico 7.

Distribuição por gênero no perfil

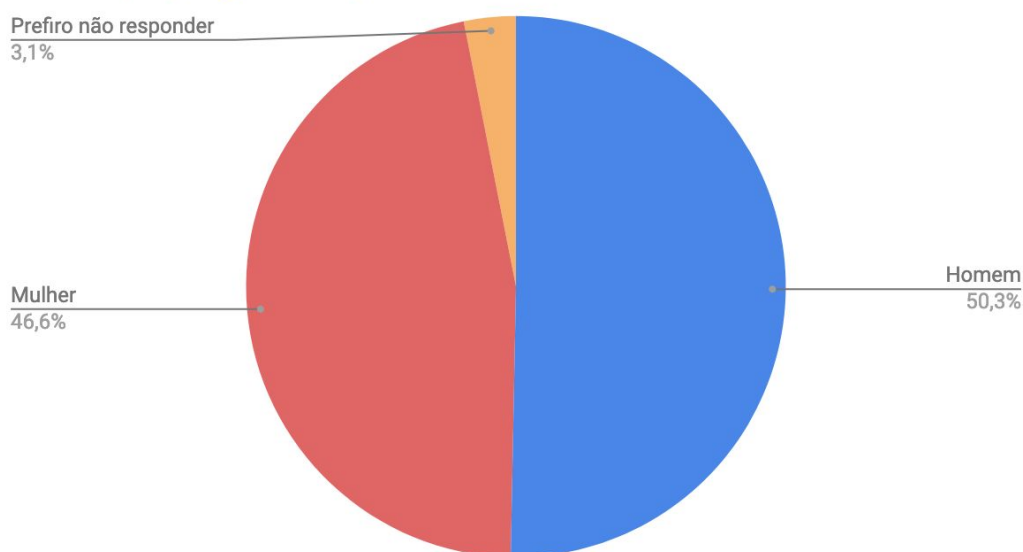


GRÁFICO 7 - Declaração de gênero dos respondentes do Rio de Janeiro com mais de 18 anos

Os principais meios de acesso à internet deles eram o celular, o *laptop* e o computador *desktop*, como é demonstrado no gráfico 8. 10 respondentes do perfil declararam *smartspeakers* (Google Home e Amazon Echo) como dispositivos de acesso à internet.

Dispositivos de acesso à internet

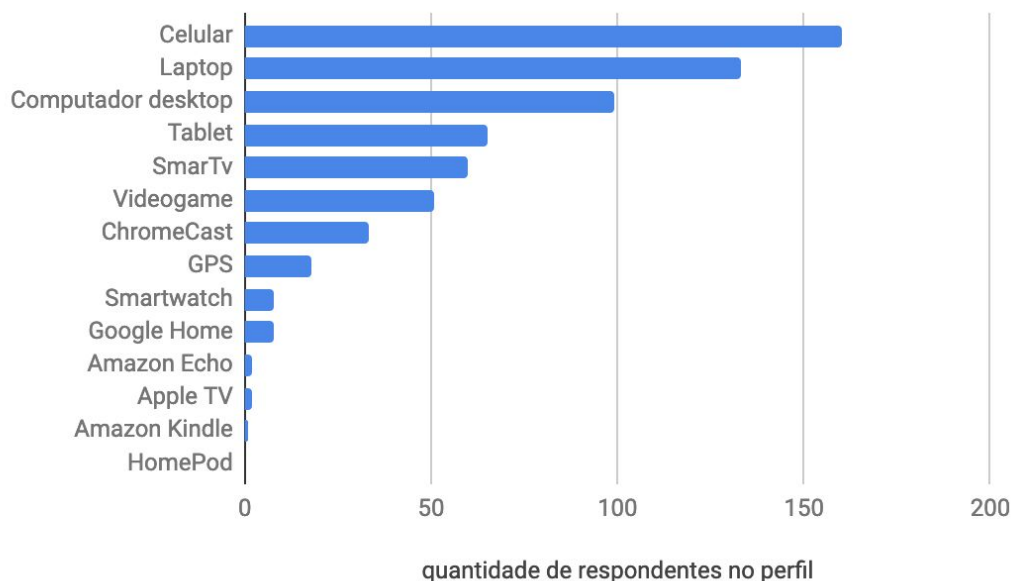


GRÁFICO 8 - Dispositivos utilizados para acessar a *internet* pelos respondentes maiores de idade do Rio de Janeiro

Por fim, a maioria (63,4%) do perfil disse ter usado um assistente virtual nos 3 meses anteriores ao questionário. No entanto, o número de pessoas que não teve contato com esses assistentes e não soube responder é alto, como exposto no gráfico 9. Para responder essa pergunta, os respondentes foram instruídos a considerar como assistentes virtuais os sistemas que mimetizam a interação com uma pessoa, como assistentes virtuais de *telemarketing*, Siri, Cortana, Google Assistente, Bixby, Alexa, e assistentes virtuais que respondem via chat (*chatbots*).

Uso de assistentes virtuais nos 3 meses anteriores

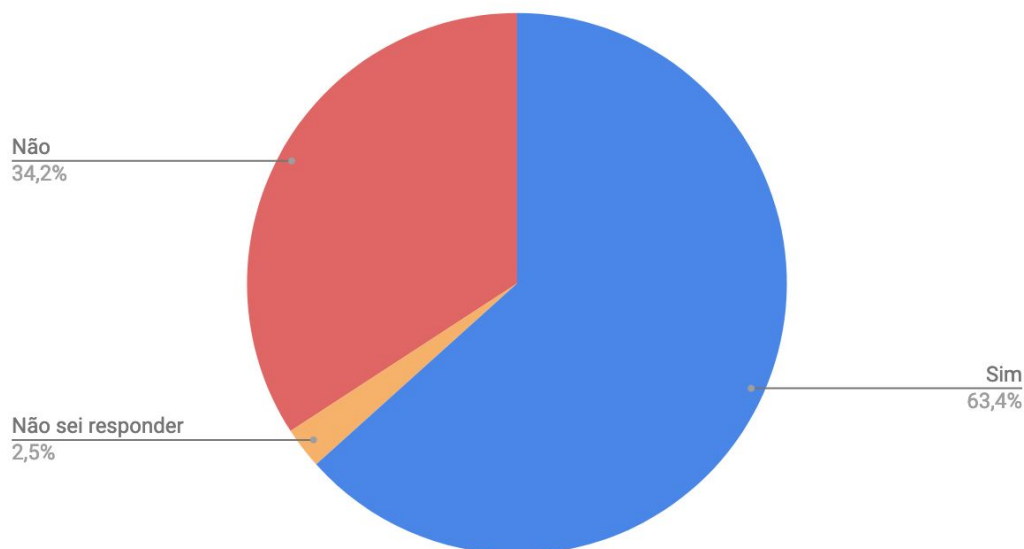


GRÁFICO 9 - Uso recente de assistentes virtuais por respondentes maiores de idade residentes no Rio de Janeiro

4.1.3 Sobre o uso dos assistentes pelo perfil selecionado

Visando entender melhor a adoção e a experiência de uso dos assistentes, trataremos os próximos dados filtrados apenas pelos respondentes que tiveram contato recente com os mesmos. Não é possível avaliar uma experiência que não se teve, e é mais confiável entender a possível adoção de um produto através de quem já o usou. A janela de 3 meses é feita para garantir que a memória de uso ainda está fresca, e que a interação foi com as últimas versões dos assistentes.

"Não se pode avaliar uma inovação ao pedir a clientes em potencial que dêem suas opiniões. Isso exige que as pessoas imaginem alguma coisa com a qual elas não têm nenhuma experiência. Suas respostas, ao longo da história, têm sido notoriamente ruins. As pessoas disseram que gostariam realmente de ter alguns produtos que, quando lançados no mercado, fracassaram. De maneira semelhante, disseram que simplesmente não estavam interessadas em produtos que depois se tornaram gigantescos sucessos comerciais." (NORMAN, 2008, p.94)

A primeira pergunta sobre o uso dos assistentes levantou quais mais haviam sido utilizados no período recente (3 meses anteriores). Os mais citados foram a

Google Assistente, os assistentes de *telemarketing* ou telefonia, os *chatbots* de *sites* e aplicativos, e a Siri, conforme mostra o gráfico 10.

Assistentes virtuais com uso recente declarado

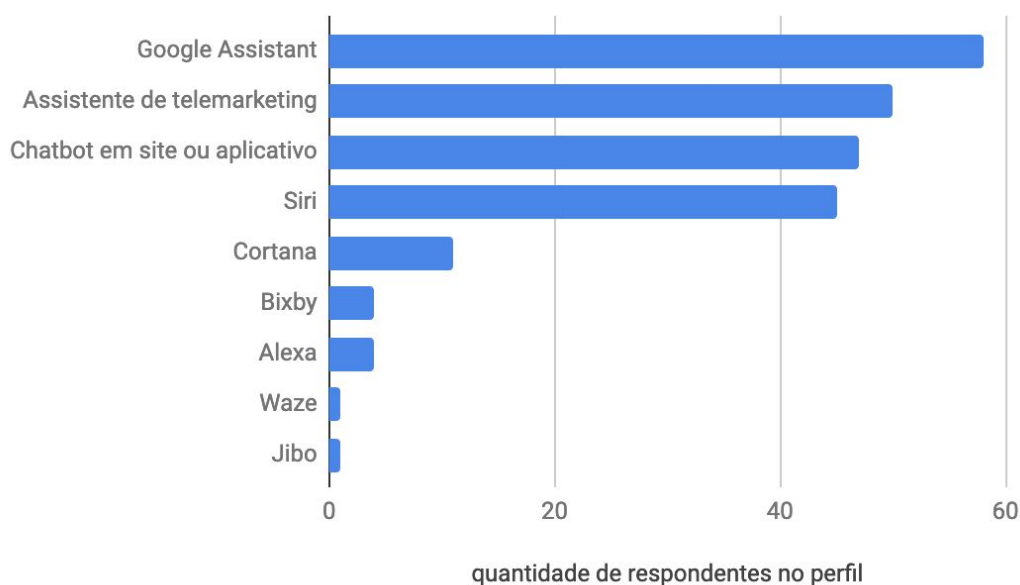


GRÁFICO 10 - Assistentes mais usados pelos respondentes do perfil

É interessante notar que, mesmo os assistentes virtuais que não utilizam interfaces conversacionais e que não possuem uma personalidade desenvolvida (não sendo considerados AVPs para essa pesquisa), foram lembrados como assistentes que mimetizam uma pessoa. Essa descoberta dialoga com Nass e Brave (NASS; BRAVE, 2007), que ressaltam a atribuição automática de uma personalidade, e por tanto de uma pessoa a uma voz.

"O cérebro humano raramente faz distinção entre falar com uma máquina e com uma pessoa - mesmo com máquinas de qualidade muito ruim na compreensão e produção de fala."(NASS; BRAVE, 2007, n.p, tradução nossa)

Dentre os dispositivos usados para acessar esses assistentes, o celular é o que mais se destaca com 93,1% dos respondentes tendo o declarado. O grande uso por esse meio reforça a importância do mesmo para tornar os assistentes e as

interfaces de voz acessíveis ao público, como previsto. A lista completa dos dispositivos utilizados pode ser consultada no gráfico 11.

Dispositivos usados para acessar o assistente

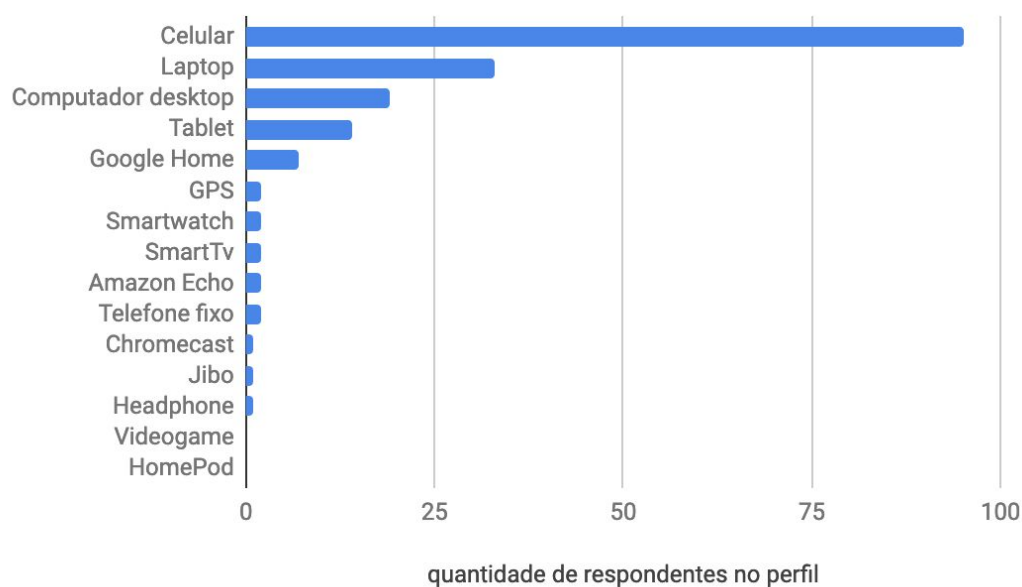


GRÁFICO 11 - Meios de acesso aos assistentes

As finalidades de uso mais citadas foram: fazer uma pergunta ou pesquisa, curiosidade (explorar o assistente) e saber sobre a previsão do tempo. A lista completa com os propósitos de uso pode ser consultada no gráfico 12.

Finalidades de uso do assistente



GRÁFICO 12 - Principais objetivos de uso dos assistentes

Os respondentes também avaliaram os assistentes que usaram de acordo com a experiência e a adoção. Quanto à experiência, 54,7% classificou como boa, 20,6% avaliou como indiferente e 14,7% considerou a experiência ruim. A maioria (67,6%) continua usando o assistente. A experiência e adoção dos respondentes do perfil pode ser consultada no gráfico 13.

Experiência e adoção dos assistentes (no perfil)

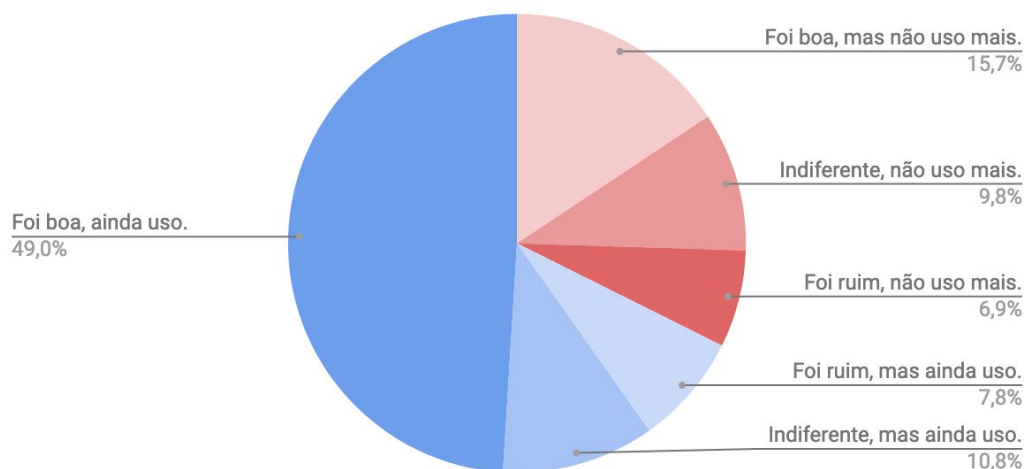


GRÁFICO 13 - Experiência e adoção dos assistentes

4.1.4 Conclusões sobre o questionário

A adoção dos assistentes é baixa (42,9%) quando levamos em consideração todos os respondentes maiores de idade do Rio de Janeiro, apesar de ser alta entre os que tiveram contato recente com os mesmos, conforme mostra o gráfico 14. Esse dado pode indicar que ainda há pouco conhecimento sobre os assistentes.

Adoção dos assistentes

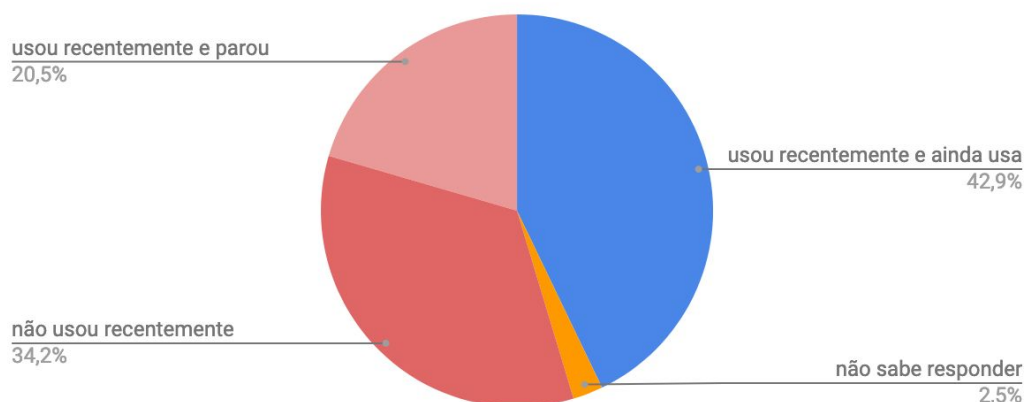


GRÁFICO 14 - Adoção dos assistentes entre que usou ou não nos 3 meses anteriores

A adoção difere quando observada por assistente. Dentre os assistentes mais usados, os AVPs - que utilizam interfaces conversacionais e que possuem uma personalidade desenvolvida - foram os com maior grau de adoção, podendo indicar que a premissa da pesquisa é verdadeira. (gráfico 15)

Adoção dos assistentes mais usados

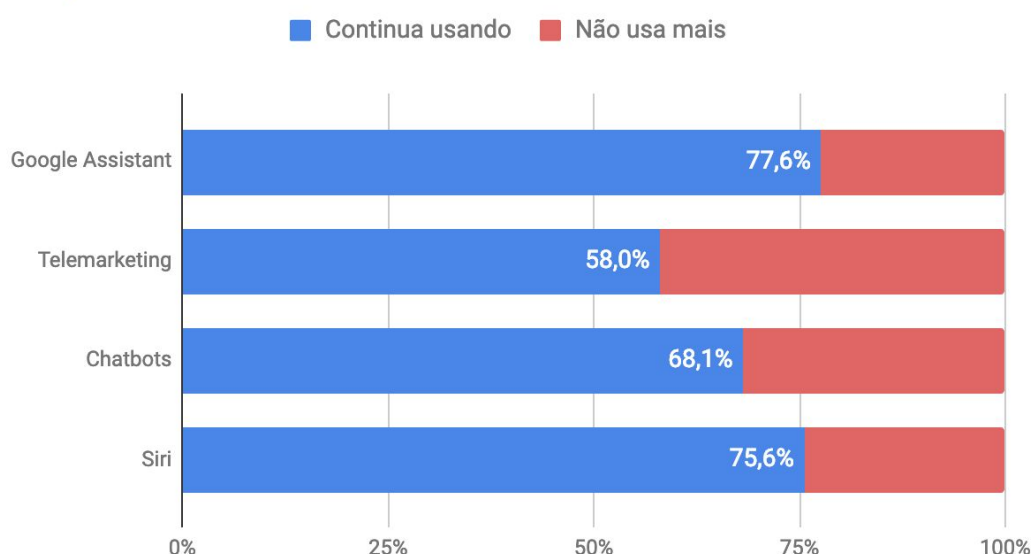


GRÁFICO 15 - Grau de adoção dos residentes do Rio de Janeiro maiores de idade dividido por assistente

O melhor avaliado quanto à experiência de uso é a Google Assistente, como ilustrado no gráfico 16. A Siri, no entanto, é a pior avaliada, apesar do alto grau de adoção. Um pressuposto que pode se formar é que, apesar das interfaces conversacionais e do desenvolvimento de uma personalidade ajudarem na adoção, para que a experiência seja avaliada positivamente, a personalidade deve ser adequada ao usuário do assistente, como foi sugerido por Nass e Brave (NASS; BRAVE, 2007). Caso essa indicação seja verdadeira, a personalidade mais introvertida e amigável do Google Assistente pode estar sendo mais bem avaliada que a personalidade extrovertida e sarcástica da Siri, por conta da personalidade dos usuários respondentes. Essa possibilidade será discutida mais a frente, no resultado dos grupos de foco.

"A identificação social - decidir se alguém 'é como eu' ou 'não é como eu' - é crítica para um princípio humano fundamental: quanto mais semelhantes forem duas pessoas, mais positivamente elas estarão dispostas em relação uma a outra." (NASS; BRAVE, 2007, n.p, tradução nossa)

Avaliação da experiência dos assistentes mais usados

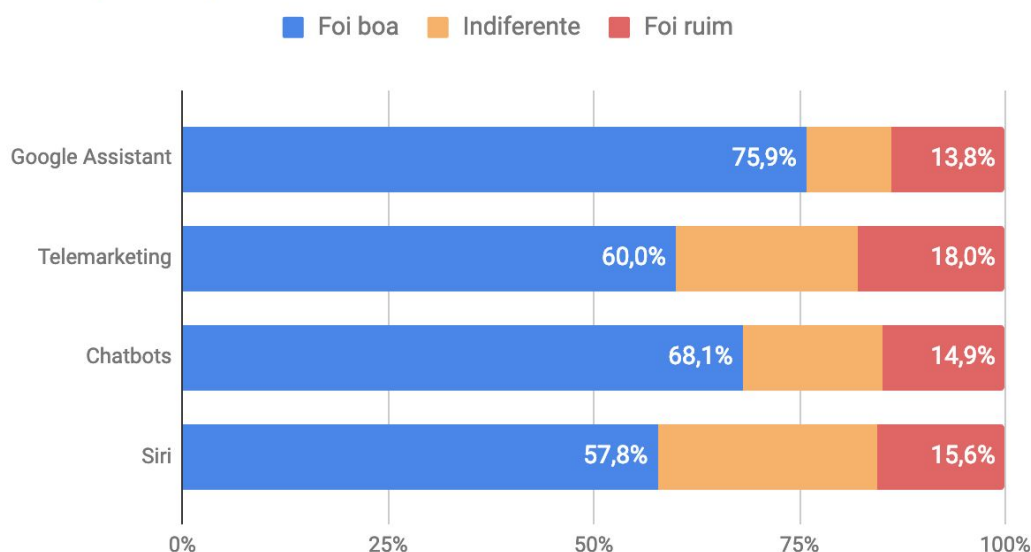


GRÁFICO 16 - Avaliação da experiência de uso por assistente

Um dos respondentes comentou sobre sua experiência de uso, reforçando a importância das interfaces conversacionais com personalidade desenvolvida: "Acho que na pergunta acima [sobre experiência de uso] depende: quando é de *telemarketing*, considero indiferente ou ruim. Já outros, como do Google, Siri ou Cortana, foi boa." Esse respondente classificou sua experiência de uso com assistentes como boa, mas não continua usando-os.

Quanto ao uso da voz como interface, houveram comentários diferentes. Alguns respondentes demonstraram reconhecer vantagens no uso das interfaces de voz, tendo citado principalmente a preguiça como motivador:

"Apesar da minha experiência não ser muito boa quando uso assistentes de voz, acho que o principal motivo que me faz continuar tentando é a preguiça, principalmente quando estou deitada ou não quero ficar olhando pra uma tela brilhando na minha frente. Isso me leva a perguntar 'Hey, Siri, que horas são?', ou

'Hey, Siri, como está o tempo hoje?'. ' (Respondente que avaliou sua experiência de uso como ruim, mas que continua usando os assistentes)

Além da preguiça, apareceu também a preocupação com a segurança ao usar o celular na rua. Esse aspecto não foi observado durante os estudos da revisão bibliográfica, potencialmente sendo específico do público estudado:

"Costumo usar quando a pergunta é grande e tenho preguiça de escrever. Na rua costumo usar como forma de evitar mexer no celular e não ser roubado. " (Respondente que avaliou sua experiência de uso como boa, e continua usando os assistentes)

Outros respondentes deram mais peso às desvantagens do uso da voz como interface. Um dos fatores levantados foi o compartilhamento de informações particulares: "Prefiro inserir dados pela tela. Tenho mais privacidade". Outra desvantagem é a percepção de que a velocidade para obter informações é maior ao digitar na tela: "Penso que são importantes, contudo penso que é possível obter as mesmas informações e com maior velocidade com os dedos."

Por fim, houveram reclamações quanto à interpretação errada de pedidos do usuário, levando a frustração: "Gosto muito da Siri, porém me incomoda quando pergunto uma coisa e ela me responde algo completamente diferente.". Outro ponto de atenção foi a falta de informação quanto às possibilidades do sistema, que leva a uma dificuldade no uso:

"O assistente que mais tive dificuldade foi a Cortana, no Windows. O assistente não entendia minhas perguntas e eu fiquei confusa quanto às suas funções, não consegui entender o que ela poderia fazer por mim." (Respondente que havia usado chatbots, Google Assistant e Cortana, que avaliou sua experiência de uso como ruim, e não usa mais os assistentes)

4.2 Grupo de foco

Segundo Galego e Gomes (GALEGO; GOMES, 2005), o *focus group* encontra suas raízes históricas em Merton, que defendia que as perguntas fechadas das entrevistas nem sempre proporcionavam as respostas mais exatas. Ele dizia que os resultados das pesquisas podiam ser influenciados, tanto por descuido

quanto por omissão, por conta da construção dos questionários. Além disso, o entrevistado era limitado nas suas respostas, por causa da sequência das perguntas e das respostas fechadas. Na tentativa de resolver esse problema, ele começou a desenvolver, na década de 30, um sistema que consistia em pedir para uma audiência descrever as suas respostas visando que isso trouxesse novas linhas de reflexão.

Paul Lazarsfeld, então, em 1941, convida Merton a aplicar o primeiro *focus group* para ajudar em sua tabela de respostas da audiência de programas de rádio. Lazarsfeld reúne um grupo de pessoas que refletia a típica audiência da época e lhes dá alguns botões de diferentes cores para apertar, caso achassem alguns aspectos positivos ou negativos. Após cada sessão, Merton reunia os participantes para perguntar em detalhes os motivos de suas respostas. Apesar de ter começado com programas de rádio, a técnica depois se expandiu para ser utilizada em diversos outros cenários. (GALEGO; GOMES, 2005)

O grupo de foco é uma técnica qualitativa que reúne um grupo de participantes para debater sobre um tema específico. Essa técnica não permite resultados em tabelas ou porcentagens, ao invés disso, tende a ser mais exploratória e compreender ideias mais gerais e subjetivas. A técnica tem como objetivo prover entendimento sobre as percepções, sentimentos, atitudes e motivações sobre um tópico ou produto.

Geralmente, são usados grupos de 8 a 10 participantes, pois grupos maiores são difíceis de se administrar; as interações entre os participantes acabam sendo menos efetivas e as discussões são complicadas de controlar. No entanto, existem formatos de grupos de foco com menos de 8 participantes.

Pesquisas qualitativas podem ser generalizadas para um público específico por serem baseadas em uma amostragem da população alvo. No entanto, os resultados do grupo de foco não são quantificáveis, pois apenas se aplica a um pequeno segmento do público alvo em questão, não podendo ser generalizados para a população em geral. Em contrapartida, o que o *focus group* oferece é um

entendimento mais aprofundado das opiniões e perspectivas, que uma entrevista ou formulário poderia obter.

As hipóteses geradas por um grupo de foco frequentemente levam a um outro estudo mais quantitativo, de forma a validar o que foi levantado. O oposto também é válido, sendo grupos de foco usados para interpretar mais a fundo os resultados de um estudo qualitativo.

Visando levantar a aceitação e as opiniões do residente do Rio de Janeiro sobre os Assistentes Virtuais Personificados, bem como mapear suas dificuldades com a tecnologia e sua visão das possibilidades da mesma para o futuro, a técnica de grupo de foco foi escolhida para interpretar em maior profundidade os dados obtidos pelo questionário.

4.2.1 Sobre a realização do grupo de foco

No total, foram realizados um teste piloto e dois grupos de foco com participantes de perfis e experiências variados, que haviam demonstrado interesse em participar da pesquisa ao responderem o questionário.

O grupo piloto ocorreu na PUC-Rio no dia 05 de junho de 2018 e contou com a participação de 6 pessoas. Todos eram homens com idade entre 20 e 25 anos, 3 com ensino superior incompleto, 2 com ensino superior completo e 1 com pós-graduação ou especialização completa. Para usar a internet, todos usavam celular, a maioria usava computador *desktop* ou *laptop* ou ambos e um possuía um *smartspeaker* Google Home em casa.

O primeiro grupo piloto foi realizado no espaço de *coworking* Tribo em Copacabana no dia 08 de junho de 2018 e teve 6 participantes, sendo 5 homens e 1 mulher. A idade do grupo variava de 23 a 34 anos, e 2 possuíam ensino superior incompleto, 1 tinha ensino superior completo e 3 pós-graduação ou especialização completa. Para usar a internet, todos usavam celular, todos usavam computador

desktop ou *laptop* ou ambos, e apenas um não usava *videogame*. Quanto ao uso de AVPs, 1 pessoa havia tido contato com a Cortana, 2 com a Siri, 3 com a Google Assistente e apenas 1 pessoa não havia tido contato com nenhum, apenas assistentes virtuais de *Telemarketing*.

O segundo grupo aconteceu na PUC-Rio no dia 19 de setembro de 2018, e 7 pessoas participaram, sendo 5 mulheres e 2 homens. A idade do grupo variava de 20 a 26 anos, sendo que 6 tinham ensino superior incompleto e 1 tinha ensino superior completo. Para usar a internet, todos usavam celular, todos usavam computador *desktop* ou *laptop* ou ambos, e 1 usava *videogame*. Quanto ao uso de AVPs, 3 pessoas haviam tido contato com a Siri e 5 com a Google Assistente.

Todos os participantes foram instruídos a testar algumas perguntas no assistente disponível em seu celular uma semana antes do grupo de foco, para que todos chegassem com um conhecimento comum e memória recente para a discussão. As perguntas indicadas foram:

- **Quem é você?** Aqui, o participante teria contato com a personalidade do AVP e com a forma a qual o mesmo se apresenta. Essa pergunta representa também a finalidade de curiosidade ou exploração do assistente, retratada no questionário realizado.
- **Quais livros Stephen King escreveu?** Essa pergunta mesclava duas línguas em uma mesma frase, para que o participante tivesse a experiência de tratamento do AVP nessa situação.
- **Como está a previsão do tempo para hoje?** Pergunta de aspecto funcional mais citada no questionário depois de "fazer uma pergunta ou pesquisa".
- **Ligue para minha(meu) mãe(pai), Mande uma mensagem para meu(minha) namorado(a) e Como está o trânsito até o trabalho?** Essa pergunta visava a exploração do tratamento do assistente para termos referenciados e não descritos objetivamente (por exemplo, usar "minha mãe" ao invés do nome dela). Além disso, tinha o intento de incitar a curiosidade

dos participantes sobre que informações pessoais o assistente conseguiria descobrir ou sugerir.

- **Como faço para ir até a PUC-Rio?** Pergunta funcional onde a referência está na localização do usuário, para explorar o uso de sensores e informações em tempo real.

Cada sessão foi dividida em dois blocos onde: no primeiro, o grupo discutia sobre o uso do assistente virtual e experiências passadas com eles; e no segundo, o grupo debatia sobre suas expectativas e receios para o futuro dos assistentes virtuais. Como o propósito dessa pesquisa foi exploratório, os achados da pesquisa foram divididos por grandes categorias de discussões que os grupos trouxeram e são apresentados em conclusões comparando o que os participantes apontaram e o que a revisão bibliográfica mostrou.

4.2.2 Os problemas tecnológicos

O primeiro grande tema que surgiu em ambos os grupos foram os problemas técnicos que os assistentes ainda precisam corrigir.

Os assistentes que podem ser acionados por meio da voz possuem uma frase de invocação que permite que eles reconheçam a intenção do usuário de acessá-los. Essa frase é necessária, pois, sem ela, esses sistemas não saberiam quando o usuário está falando com outras pessoas, e quando está se direcionando a eles, de forma a poderem identificar quando escutar ativamente e processar o que foi pedido. Quando os sistemas são acionados por voz, por tanto, o usuário deve esperar a confirmação que foi ativado e está escutando, antes de se direcionar a ele. Um usuário de *iOS* do primeiro grupo relatou um grande incômodo com essa demora no tempo de ativação da Siri, pois faz com que **não seja possível invocar e realizar um pedido imediatamente após**.

"Quando você fala 'E aí, Siri', esse tempozinho que demora para ela processar o 'E aí, Siri' e ativar é muito... Você nunca sabe direito quando que você pode continuar

falando, porque eu não posso falar 'E aí, Siri, me acorde 7 da manhã'. Se eu falar direto, ela não entende".

Outros participantes do grupo não tinham o mesmo incômodo, e acreditavam que a demora na ativação fosse uma questão de performance do celular ou sinal de *internet* do participante. Os participantes do segundo grupo de ambos os sistemas operacionais disseram achar ruim **precisar de *internet* para acessar o assistente**. De qualquer maneira, a frustração desse usuário pode ser atribuída ao fato da Siri não possuir um *feedback* claro quando está sem acesso a *internet* e, por tanto, não deixar claro quando pode responder. Esse assunto será tratado no segundo tópico, usabilidade.

Durante a discussão, foi forte em ambos os grupos a percepção de que, quando um assistente não consegue responder a uma solicitação, essa limitação é atribuída a um **problema de interpretação**. As pessoas levantaram que os assistentes conseguiam, em muitos dos casos, entender o que havia sido dito palavra por palavra, porém não sabiam qual ação deviam performar.

Apesar disso, alguns problemas são sim atribuídos ao não entendimento do assistente. No grupo um, os participantes reforçaram que eles são mais assertivos quando o ambiente é livre de ruídos; em ambientes barulhentos, ou quando o usuário está em movimento, existe dificuldade de captação e o assistente apresenta mais erros. Outro momento relatado de ocorrer erros no entendimento é quando mais de uma língua é falada na mesma frase.

Em ambos os grupos houveram reclamações quanto ao **entendimento de termos de outra língua** e quanto a misturar palavras de diferentes línguas em uma mesma frase, como no exercício solicitado (livros de Stephen King), ou, no exemplo trazido por um participante do segundo grupo, como ao pedir por uma música de uma banda estrangeira. No segundo grupo, os participantes chegaram a falar que o assistente é melhor quando está em inglês.

Outro problema técnico relatado foi a forma de acessar aplicativos quando se usa o assistente por meio da fala, pois ele **não te permite continuar**

interagindo com a voz dentro de aplicações terceiras. O assistente apenas abre o aplicativo e, a partir daí, o usuário deve proceder pela interface visual. Apenas abrir um aplicativo com a voz foi visto como uma funcionalidade "inútil".

No segundo grupo, um participante usuário de *iOS*, que também já havia usado a Google Assistente, disse não gostar da forma que o *Google* trata aplicativos terceiros que suportam o uso da voz. Por levar o usuário a falar com outra entidade, o usuário disse que o sistema parece estar te transferindo, e não gosta disso. Os outros participantes ainda não haviam experimentado essa funcionalidade.

Outra percepção que foi compartilhada pelo segundo grupo, mas que não foi mencionada no primeiro, é sobre a limitação dos assistentes hoje. Os participantes acreditam que **existe a capacidade de os assistentes fazerem muito mais pela quantidade de informações que eles captam, mas não o fazem.** Um participante até disse acreditar que a Apple impõem uma limitação para a Siri para que ela não entre nos aplicativos e faça dadas tarefas. A crença de que a Siri é mais limitada que a Google Assistente foi percebida em ambos os grupos, e é uma razão para a avaliação pior de sua experiência de uso.

Ambos os grupos acreditam que **as limitações técnicas serão superadas**, pois são problemas que a evolução da tecnologia pode resolver. O segundo grupo ainda chegou a conclusão que, demorando mais ou menos, os assistentes de voz vão acabar sendo lugar comum, pois as pessoas acham divertido e não há nenhum impeditivo tecnológico.

4.2.3 Usabilidade e eficiência em assistentes

As pessoas perceberam o uso da voz e dos assistentes como uma forma eficiente de **encurtar caminhos de navegação na interface do celular.** As atividades que dependem de muitos passos, como criar um evento em uma agenda

ou ligar para um conhecido, conseguem ser realizadas muito mais rapidamente, como ilustrado a seguir:

- **Criar um evento na agenda sem o uso de assistente** requer 9 passos. Deve-se destravar o celular, abrir o aplicativo de agenda, clicar para criar um novo evento, dar um nome, marcar uma data, marcar um horário, encontrar os contatos que pretende chamar e adicioná-los. Enfim, concluir a criação do evento e enviar os convites.
- **Criar um evento na agenda com o uso de assistente** requer 3 passos. Deve-se destravar o celular, invocar o assistente e falar, por exemplo, "marcar reunião de planejamento com A, B e C na quarta-feira, às 15 horas".

Apesar disso, o segundo grupo retratou que a necessidade de destravar o celular através da tela faz perder o sentido de usar o assistente por meio da voz. Isso se dá principalmente em situações onde o usuário não pode usar as mãos, como é o caso da direção de um carro.

Ainda assim, mesmo quando não se está impossibilitado de manusear o celular, **as pessoas esperam que o assistente não dependa da seleção de opções na tela após uma interação por voz**. Esse tipo de tratamento ocorre, por exemplo, no caso de desambiguação, quando se pede para ligar para "Maria", quando há mais de um contato registrado com esse nome no celular, ou quando se pede as direções para um restaurante sem especificar a filial.

A interface multimodal⁷, que utiliza mais de uma modalidade de interação (voz e texto, voz e gráfico, etc) não é percebida de maneira negativa, no entanto, as pessoas demonstraram presar pela consistência: ao acessar o assistente usando a tela, a tela pode servir para complementar a experiência; ao acessar o assistente usando apenas a voz, a interface deveria estar preparada para responder apenas pela voz.

⁷ O conceito de interface multimodal foi apresentado no capítulo 3, página 50.

Quanto à invocação, houveram divergências de preferência. Enquanto alguns gostavam de ativar o assistente pela voz, alegando que precisar de um botão era "chato", um caminho mais longo; outros preferiam o botão, pois assim não corriam o risco de ativar o assistente sem querer. No segundo grupo, quase todos preferiam a ativação por voz, com exceção de situações onde se sentiam expostos, como interação em ambientes públicos. Essa descoberta corrobora com o que foi apontado por Weiser (1994), quando dizia que **a interação por voz em um ambiente com outras pessoas não seria bem-vinda**. Cathy Pearl (2016) também chegou a mesma conclusão:

"Hoje, muitos de nós trabalham em escritórios de espaço aberto. Imagine pedir para o seu computador para realizar tarefas: 'Computador, encontre todos os meus documentos *Word* dessa semana'. Agora, imagine todos no escritório fazendo isso! Seria um caos. Além do mais, quando você fala, qual dos computadores está escutando?" (PEARL, 2016, n.p, tradução nossa)

Um outro tópico relevante quanto a esse assunto, foi o **incômodo com a necessidade de repetir a invocação durante uma mesma conversa** com o assistente. No segundo grupo de foco, os participantes disseram preferir que a conversa ficasse ativa até fechar o aplicativo ou até solicitar que encerrasse. No entanto, um participante levantou que a interação contínua poderia ser um problema quando não se está totalmente dedicado ao assistente.

O tempo que se deve esperar para precisar de uma nova invocação não é fácil de ser definido, mas a Cathy Pearl (2016), sugere o uso de 10 segundos:

"A janela de tempo na qual seu sistema deveria continuar ouvindo após a palavra ou o clique no botão de invocação deve ser escolhido com cuidado. Sendo muito curto, você pode perder usuários que hesitaram brevemente antes de falar; sendo muito longo, o sistema poderá estar ouvindo uma conversa que não é destinada a ele. Como um princípio básico, 10 segundos é um bom ponto de partida." (PEARL, 2016, n.p, tradução nossa)

Ademais, a invocação da Siri também foi um ponto de discussão, como foi pontuado no primeiro tópico. Os usuários de *iOS* do segundo grupo de foco também demonstraram desgosto por não saberem se a Siri está ouvindo ou não quando ficam sem *internet* ou com a *internet* lenta.

A **falta de *feedback* adequado** pode estar relacionado com o ponto discutido no capítulo 2: novas formas não visuais de retorno para o usuário precisam ser pensadas para esse novo cenário. Esse problema de usabilidade do sistema é uma das razões para a experiência de uso não ter sido avaliada de forma positiva no questionário.

Além da invocação, uma reclamação realizada sobre o assistente da *Apple* foi perder o contexto do diálogo. A percepção que os seus usuários tiveram é de que toda frase parece ser uma nova conversa, pois o **assistente não resgata informações** ou se lembra do assunto sobre o qual estava sendo conversado. Esse é um ponto de atenção, pois a interface conversacional da Siri ainda não está bem desenvolvida, agindo por vezes como uma interface de comando e controle de voz. A avaliação menos positiva sobre a experiência de uso da Siri e a reclamação sobre a interação não responder de acordo com o esperado são indicativos de que o uso de interfaces conversacionais de fato colaboram para uma percepção positiva sobre o assistente.

Um último ponto de forte atenção, pois pode desencorajar o uso das interfaces de voz é, ao responder o usuário, o assistente **oferecer uma resposta muito grande a uma pergunta simples**. Quando isso acontece, o usuário deixa de usar a interface de voz para olhar a resposta rapidamente no celular, ou se condiciona a mudar a forma que geralmente perguntaria na tentativa de obter uma resposta mais específica.

4.2.4 Os momentos e usos dos assistentes

No uso geral, os dois grupos **não sabiam dizer ao certo o que os assistentes eram capazes de fazer**. Isso se mostrou poder ser, ao mesmo tempo, bom e ruim. Alguns participantes demonstraram gostar de testar e ir descobrindo do que o assistente é capaz, pois se divertem explorando o assistente e encontrando novas funcionalidades.

No entanto, apesar da grande satisfação quando se testa um pedido e o assistente funciona, uma única experiência ruim pode acabar com a confiança que o usuário tem naquele sistema. Um dos participantes do primeiro grupo exemplificou essa situação quando tentou impressionar a namorada, e o sistema não respondeu de maneira adequada, culminando no abandono do assistente.

O participante atribuiu o problema ocorrido a falta de *feedback* sonoro da Siri quando o sinal de *internet* está fraco, ou quando se está sem internet. O usuário invoca o assistente que, por vezes demora, e por vezes não atende, sem explicar para o usuário se ele está sendo escutado ou não.

Além disso, alguns dos usuário do segundo grupo demonstraram não gostar de explorar o assistente e, dessa forma, tentam se adequar às limitações dos mesmos. Após algumas tentativas frustradas, as pessoas passam a pedir apenas o que sabem que o assistente é capaz de fazer e experimentam menos. Isso foi algo notado tanto em usuários da Siri (*Apple*) quanto da Google Assistente (*Google*).

Os assistentes também são percebidos pela maioria como **uma forma de entretenimento, e não uma ferramenta**, e dentre os que consideram isso, a frequência de uso é baixa. Ainda assim, mesmo quem o percebe como ferramenta, acredita que eles estão sendo desenvolvidos mais por serem divertidos que por serem úteis. Além disso, os participantes que relataram usar mais esses sistemas possuíam um *Google Home*, e atribuem sua frequência de uso a essa possibilidade de usar o assistente dentro de casa no seu cotidiano.

Apesar disso, o primeiro grupo levantou que faz parte do ciclo das novas tecnologias começarem como entretenimento e depois evoluírem para outros usos. O grupo exemplificou dizendo que as ferramentas de *chat* começaram sendo usadas para bate-papo, e que hoje são usadas em empresas e no atendimento de clientes, fazendo parte do dia-a-dia. Para eles, os assistentes estão na fase "*bate-papo da UOL*" da adoção.

As razões mais citadas para usar os assistentes foram a preguiça e o entretenimento. Com relação a preguiça, os participantes relataram usar quando estão indo dormir e não querem usar as mãos para configurar um despertador, quando precisam de uma informação mas não querem parar o que estão fazendo para buscar ou ainda para encurtar caminhos de navegação nas interfaces dos celulares (por exemplo, perguntar a previsão do tempo para uma viagem, procurar um contato para ligar ou marcar um evento na agenda).

O uso por entretenimento consiste em testar funcionalidades por curiosidade, conversar com o assistente para ver as brincadeiras que fala ou ainda encontrar *easter eggs* (segredos escondidos na aplicação que possuem caráter humorístico).

4.2.5 Acessibilidade

Apesar de não haver nenhum deficiente visual nos dois grupos, em ambos surgiu a crença de que **os assistentes serão muito úteis para quem não consegue usar interfaces gráficas.** A humanização trazida pela persona que os assistentes adotam foi vista como uma boa forma de tratar as interfaces de voz para que fiquem mais confortáveis de usar para quem precisa delas.

Nos dois grupos, os participantes disseram acreditar que os assistentes irão facilitar o uso para a acessibilidade, pois julgam que falar o que se quer é mais rápido que tentar se encontrar em uma interface gráfica usando a audição.

Apesar disso, um dos participantes levantou que a questão da acessibilidade só está sendo discutida junto com os assistentes por acaso. Ele não acredita que esses sistemas foram pensados para serem acessíveis, é apenas conveniente que sejam, e então se torna um argumento de venda.

4.2.6 Os assistentes no ambiente conectado

Foi comum a percepção de benefício de um cenário conectado, ambos os grupos mostraram acreditar que o futuro dos assistentes é não estarem mais presos aos *smartphones*. No primeiro grupo, o destaque foi para o ambiente doméstico, onde não há tanto ruído quanto na rua, e onde se tem mais liberdade. O segundo grupo focou mais na possibilidade de futuras integrações, como com a realidade aumentada e novas tecnologias; suas referências eram da ficção, sendo muito citado o assistente Jarvis do filme Homem de Ferro (FAVREAU, 2008).

A convicção comum é que **os assistentes serão mais úteis no ambiente conectado**. O benefício claro é substituir o que hoje é feito de forma mecânica, sem que ninguém precise tocar em quase nada: os assistentes saberiam o que precisa ser comprado para a despensa, controlariam as luzes, ajustariam a temperatura do ar condicionado, etc. No entanto, não foram todos os participantes que gostaram dessa ideia, muitos se mostraram incomodados com a ideia de um sistema tão vinculado às suas vidas.

No segundo grupo, foi levantado que seria interessante romper a barreira física dos objetos. Um dos participantes chegou a discutir que as interfaces são muito visuais, e seria bom que as aplicações pudessem se comunicar de forma mais natural com audição e outros sentidos. Para ele, a integração dos assistentes com o mundo conectado pode gerar novas formas de interação, o que empolgou também os outros participantes. Ainda assim, muitos demonstraram preocupação com a privacidade das pessoas nesse contexto.

Os participantes chegaram a conclusão que, quando tudo está integrado, é ainda mais fácil saber tudo sobre uma pessoa. Isso pode tornar o sistema mais pessoal por um lado, mas por outro, a segurança, a privacidade e outras questões que discutiremos em um próximo tópico podem ser impactadas negativamente.

4.2.7 A customização e a personalização dos assistentes

Ambos os grupos consideraram essencial que **a personalidade do assistente fosse personalizada** para cada usuário, ou seja, que se adaptasse para seu tipo de humor, para suas preferências e se comunicasse da forma que é melhor pra ele. Alguns participantes trouxeram também a possibilidade do assistente se adaptar "observando" como você age com outras pessoas, e ir mudando e evoluindo, aprendendo como uma "inteligência artificial social".

Essa visão está de acordo com Nass e Brave (2007), que colocam a identificação social como ponto importante para a adoção e boa avaliação de uma interface de voz. No experimento que realizaram, chegaram a seguinte conclusão:

"Os efeitos da personalidade da voz se estenderam para os sentimentos dos usuários sobre as descrições dos livros. Os extrovertidos gostaram mais das resenhas dos livros quando lidas por uma voz extrovertida, e introvertidos gostaram mais das resenhas quando lidas por uma voz introvertida, mesmo o conteúdo de ambos sendo idêntico. De forma similar, as pessoas acharam as descrições dos livros mais confiáveis quando a voz que lia as descrições combinava com suas próprias personalidades." (NASS; BRAVE, 2007, n.p, tradução nossa)

Apesar disso, o segundo grupo se mostrou aberto para que **diferentes situações pedissem assistentes com personalidades plurais**, como por exemplo, para transparecer a visão de uma empresa ou estabelecimento no atendimento. No entanto, foi frisado que, nesse caso, o assistente não seria do usuário.

Os participantes acreditam que, para um assistente ser pessoal, ele deve conhecer bem seu usuário e ser capaz de reconhecer seu estado de humor, dessa forma adaptando sua comunicação. **Os participantes demonstraram que não gostariam de ser tratados da mesma forma em dias em que estão bem e em dias em que seu humor está abalado.** Um participante disse que, "se ela [Siri] vier ser sarcástica em um dia que eu não estou legal, eu vou chutar o celular".

Essa constatação sobre o humor também foi percebida no estudo de Nass e Brave (2007). Eles apontam que os usuários preferem quando há consistência entre estado emocional do conteúdo e da voz, estranhando se um conteúdo triste é

falado de forma alegre, ou um conteúdo empolgante é passado através de uma voz desanimada.

Mais que isso, eles demonstram que os usuários são contagiados por emoções fortes, sejam elas positivas ou negativas. Dessa forma, se uma interface de voz demonstra muita felicidade para um usuário que está com raiva, é possível que essa pessoa seja contagiada pela intensidade da emoção, mas mantenha seu humor inicial, ficando com muita raiva. (NASS; BRAVE, 20017)

A maioria dos participantes vê como positivo, no futuro, ter um assistente que conhece muito bem seu usuário a ponto de poder resolver afazeres sem que ele se preocupe ou precise tomar uma ação (como, por exemplo, agendar compromissos, pois já sabe seus horários e agendas). Em um ambiente conectado, o assistente poderia também ajustar as coisas ao gosto do usuário, moldar a luz e a temperatura para o momento que o usuário chegar em casa. No entanto, algumas pessoas demonstraram desconforto com essa ideia.

Em ambos os grupos, surgiu também a discussão sobre a **customização ativa** do usuário. O assistente deveria se adaptar ao usuário, mas também permitir a customização, desde o tom de voz à personalidade. Algumas pessoas gostariam que ele fosse mais brincalhão e simpático, outros disseram preferir que ele fosse mais direto e educado. Para os participantes, até a forma de interação deveria ser acordada entre o usuário e o assistente, cabendo a cada pessoa decidir se quer que o assistente converse por voz ou não. Isso demonstra que, mesmo quando a interface conversacional é bem aceita por uma maioria, o gosto pela mesma varia de pessoa para pessoa.

O segundo grupo também levantou que o **excesso de personalização pode colocar quem usa esses assistentes em uma bolha**, onde só se vê o que gosta e o que conhece. Foi pontuado que isso não é um problema apenas para os assistentes, pois já é uma preocupação hoje com outros sistemas. Ademais, os participantes

trouxeram a preocupação que essa personalização excessiva seja aplicada a publicidades dentro dos assistentes.

4.2.8 A humanização e o assistente personificado

Os dois grupos retrataram os atendentes virtuais de *telemarketing* como um exemplo de interação ruim com assistentes. Todos os pontos levantados diziam respeito a falta de uma interação mais natural, a uma humanização forçada e a uma demora para realizar um atendimento.

Quanto a interação, os participantes relataram que os atendentes virtuais de *telemarketing* são muito robóticos, que sua fala não é natural e que não é possível interagir com eles. É necessário esperar uma longa fala gravada e, então, digitar uma opção ou falar uma palavra específica, o que gera uma interação pouco natural.

Para a maioria, a tentativa de humanização nos assistentes virtuais de *telemarketing* não é efetiva. Os participantes contam que ficam impacientes ao precisar ouvir a mesma mensagem pausadamente várias vezes, e quando existe a possibilidade, pulam direto para o atendimento com uma pessoa real. Os esforços para criar elementos "reais" nesses assistentes como, por exemplo, simular o barulho de um teclado, não são bem vistos. Para a maioria dos participantes, **se for para simular algo que pareça real, a interação também tem que ser mais natural**. Caso contrário, fica inconsistente e gera frustração.

Por último, as pessoas apontaram que esses assistentes geram uma demora no atendimento. Como não é possível sair das opções dadas, muitos acreditam que seria mais simples e rápido poder falar com uma pessoa, e ser transferido é sempre o último passo. Ainda assim, os participantes assumiram que a percepção sobre os assistentes de *telemarketing* pode ser prejudicada, pois a interação com eles sempre ocorre em momentos de estresse.

Em todos os outros casos, **a personalidade nos assistentes foi considerada um ponto positivo para torná-los menos robóticos**, e é esperado que se veja cada vez mais isso no futuro. Foi levantado, em ambos os grupos, que um assistente não deve ser sempre igual, pois ele vai ser previsível e se tornará chato.

Em ambos os grupos, os usuários de *iOs* consideraram positivo o comportamento da Siri ao fazer piadas, ser rebelde e falar de maneira sarcástica. Acreditam que isso traz personalidade a ela, mas que poderia ser mais trabalhado com personalização, afinal "A Siri não é a minha Siri, é a mesma Siri de todo mundo".

A assistente do *Google* foi considerada mais neutra que a do *iOS* nos dois grupos. Os usuários de Android que usaram a Google Assistente se dividiram em duas opiniões:

- A assistente do Google ainda está correndo atrás de ter uma personalidade, é sempre passiva e prestativa demais. A relação que traçaram foi com os assistentes de vilões em filmes, que sempre buscam aprovação;
- A assistente do Google é legal por ter uma personalidade amigável e ser prestativa;

A diferença de percepção sobre a personalidade da Google Assistente parece se dar pela personalidade dos usuários. As pessoas que demonstraram ser mais calmas e introvertidas durante o grupo de foco mostraram maior apreço pela personalidade demonstrada por essa assistente.

Em ambos os grupos, algumas poucas pessoas demonstraram não querer assistentes personificados com a interação por voz, e a justificativa dada é que eles remeterem a mais uma pessoa com quem precisarão lidar todos os dias. As pessoas que trouxeram esse ponto transpareceram ser mais tímidas e reservadas ao longo das discussões. Essas pessoas também trouxeram que, por acreditarem que

essa é uma tecnologia que será bem aceita pela maioria e que será usada em larga escala, elas serão **"obrigadas" a usá-la para não se sentirem excluídas**.

Nos dois grupos, os participantes se referiram aos assistentes de maneira bem personificada, chamando-os de "ele", "ela" e atribuindo características humanas como "educado", "gentil", "burra" e "sarcástica". Uma das participantes do segundo grupo, inclusive, disse agradecer e pedir "por favor" ao assistente.

A falta de possibilidade de conversar em assistentes que não são personificados, como o da TV, chama a atenção dos participantes. Apesar de não ser visto como algo bom ou ruim, esses assistentes foram considerados incompletos.

Um dos usuários do primeiro grupo considerou preocupante tentar gerar uma interação humana com algo que não o é. Disse que, por estudar automação, acredita que colocar personalidade, piadas e brincadeiras no código é uma forma desnecessária de complicar o algoritmo, apesar de achar importante humanizar. A maioria nos dois grupos, no entanto, acredita que trabalhar a personalidade dos assistentes é importante para facilitar a interação e para que as pessoas usem, e é bom que as empresas estejam desenvolvendo esse aspecto.

4.2.9 A voz como forma de interação com AVPs

Apesar da curiosidade ser um dos grandes motives para a experimentação dos assistentes, não há interesse de buscar saber se todos os dispositivos da casa tem a funcionalidade de interagir com a fala; a curiosidade só se dá após a descoberta dessa possibilidade. Alguns participantes disseram também que sabem que alguns dispositivos tem a funcionalidade, mas têm preguiça de configurar.

Foi unânime a percepção de que a necessidade de destravar o celular e selecionar opções na tela, faz perder o sentido de usar um assistente por meio da voz, pois "se você está usando a voz é porque não está podendo usar a mão".

Foram citados de forma específica os casos de uso durante a direção no carro, e o uso para telefonar para alguém. Existe a expectativa de se conseguir **fazer o que se faz hoje em dia no celular sem precisar mexer em nada**, apenas com a voz.

Apesar desse gosto pela voz, alguns participantes dos dois grupos levantaram que **preferem usar os assistentes através do texto escrito**. Por usar muito o celular na rua, às vezes eles se privam de usar o assistente para que não se exponham falando alto. Ainda assim, essas pessoas gostam de pedir algo por voz para o assistente algumas vezes. Nesse caso, ao falar em voz alta na rua, os participantes demonstraram que gostariam que a resposta não fosse dada por voz.

Esse comportamento também é pontuado por Cathy Pearl: "Muitas pessoas passam horas em seus celulares, muitas das quais estão digitando. Esse é o seu modo padrão, e elas podem não querer mudar para a voz". (PEARL, 2016)

Mesmo quando não estão em público, algumas pessoas não gostam mais de precisar falar ao telefone, preferindo o atendimento em *chat*, e levam essa preferência para os assistentes. Mesmo sabendo que os assistentes em *chats* de texto são máquinas, consideram a experiência positiva, pois não se sentem "idiotas" em usar.

4.2.10 A voz dos assistentes

As vozes dos assistentes são majoritariamente femininas e os participantes acreditam que elas foram pensadas e projetadas para serem assim. Isso foi alvo de discussão e dividiu as opiniões de ambos os grupos. Enquanto para alguns, a voz feminina como padrão foi vista como **herança de uma subserviência dos papéis femininos**; para outros, ela nunca incomodou ou foi uma questão, e até preferem, pois já estão acostumados com ela.

O gênero na voz também foi tratado por Nass e Brave (2007). Em seus estudos, perceberam que também há um esforço de identificação social através do

gênero, com mulheres preferindo vozes femininas, e homens preferindo vozes masculinas. Eles também perceberam uma tendência geral a dar mais credibilidade para vozes masculinas que femininas, principalmente quando essas vozes tratam de assuntos estereotipicamente categorizados como masculinos. A esse fato, se atribui culpa aos estereótipos de gênero.

Muitos dos participantes nunca haviam se perguntado a razão da maioria feminina nas vozes; estavam acostumados, pois isso ocorre em todo lugar, mesmo fora do ambiente virtual. Os participantes do segundo grupo tentaram ainda levantar exemplos de vozes masculinas, porém só conseguiram lembrar de 2: Eduardo, o atendente virtual da Oi, e o assistente Jarvis, do filme "O Homem de Ferro" (FAVREAU, 2008).

Levantaram também que a voz masculina, por ser mais grossa, pode parecer mais rígida, enquanto a feminina, por ser mais suave, é mais confortável de escutar. Acreditam que a voz feminina é melhor por ser mais **acolhedora**, estar mais relacionada ao cuidado. Outra hipótese levantada para a voz ser majoritariamente feminina é a sua frequência, por sons mais agudos se destacarem. Esse tipo de conclusão dos participantes reforça o **estereótipo de gênero vinculado a voz**, como foi abordado por Nass e Brave (2007).

Pontuaram também que, como as vozes femininas estão mais presentes nas áreas de serviço, cria um senso comum que as relaciona com cuidado e paciência, e que se reforça, pois você sempre é atendido por mulheres. Nesse aspecto, ser atendido por um homem pode ser um pouco chocante.

Os outros participantes se manifestaram de maneira contrária a esse padrão criado, que relaciona mulheres com sensibilidade. Um participante levantou que acredita que, sendo uma tecnologia que é artificial e está sendo criada agora, é possível ter uma figura acolhedora e sensível com voz masculina. Não é necessário reforçar um estereótipo.

Um dos participantes do segundo grupo compartilhou a experiência de ter pego um *Uber* no qual o motorista reclamou da voz que lia as instruções no mapa, pois “não gostava de uma mulher mandando nele”. Isso foi percebido por todos do seu grupo como negativo, e como um estereótipo que deveria ser quebrado.

Ainda assim, a maioria prefere a voz feminina nos assistentes, independente do gênero do usuário, diferente do que foi observado por Nass e Brave (2007). É possível, por tanto, que **a preferência pela voz feminina seja uma particularidade do público estudado**. Todavia, apesar da preferência declarada, em ambos os grupos houveram manifestações de incômodo com esse padrão, e de expectativa de mudança no futuro para um cenário de vozes mais plural.

Um participante disse ter a **expectativa de que a experiência com os assistentes seja mais fragmentada e plural**, pois não estará apenas no celular pessoal, mas no trabalho, no restaurante, no contato com empresas... Da mesma forma que você vê pessoas de diferentes tipos em diferentes lugares, devemos ter diferentes assistentes. Por vezes para uma empresa se destacar pelo atendimento, ou para mostrar alguma característica dela no assistente. Esse comentário teve a adesão dos outros participantes do grupo.

Além disso, houveram mais três pontos: algumas vozes foram apontadas como "chatas", como as vozes femininas dos aplicativos de navegação *Waze* e *Google Maps*; Os assistentes em inglês foram considerados melhores por terem a voz menos “travada” que em português; e apenas um dos participantes sabia que era possível configurar a voz dos assistentes para a masculina.

4.2.11 A questão social e os impactos na vida das pessoas

No segundo grupo, uma característica que foi debatida e entendida como positiva para os assistentes de voz, bem como para outras novas tecnologias como realidade mista, foi a mudança de postura. Com os celulares, passamos a nos

fechar e olhar muito para baixo, para a tela. **Saindo da tela, a gente volta a endireitar a coluna, olhar para as pessoas e para o mundo.**

Ainda assim, **o celular onde está o assistente hoje não é visto como uma ferramenta de trabalho**, mas como uma distração. Essa ideia ainda está sendo desconstruída, mas alguns dos participantes do primeiro grupo concordaram que essa ainda pode se configurar como uma barreira para a adoção dos assistentes. Todavia, o maior empecilho discutido para a aceitação dos assistentes virtuais foi a questão de **não ser socialmente aceito conversar com o celular ou com um computador.**

A maioria não usa o assistente em público por vergonha, por acreditarem que os outros vão achar que são "malucos" ou "idiotas". Por isso, para usar os assistentes, as pessoas esperam estar em um lugar vazio. Existe a crença que ainda estamos longe de um cenário como o filme de ficção "Her" (JONZE, 2014), onde é normal que todos na rua estejam falando com um assistente.

Como pontuado por Cathy Pearl (2016), "Apesar das VUIs estarem se tornando mais comuns, não são todos que se sentem confortáveis de falar em voz alta com um computador, mesmo quando não estão em público."

Além da vergonha, foi unânime, em ambos os grupos, o desconforto das pessoas em **usar assistentes em público por se sentirem expostos**; não há privacidade em pedir algo em voz alta com outras pessoas observando. Os participantes compararam o uso de assistentes em público com as mensagens de voz no celular. Apesar de ser prático, se sentem expostos ao emitir voz e barulho em público; seja para fazer uma ligação, escutar uma mensagem de voz enviada ou utilizar um assistente.

Esse também foi um aspecto assinalado por Cathy Pearl (2016):

"Se precisarem discutir um aspecto sobre sua saúde, a maioria dos usuários não irá querer fazê-lo conversando ao telefone no trem a caminho do trabalho. E não é apenas a privacidade pelo que os usuários contam aos sistemas, são as potenciais violações de privacidade quando um VUI automaticamente lê suas mensagens de

texto em voz alta, ou avisa que é a hora de tomar uma dada medicação." (PEARL, 2016, n.p, tradução nossa)

Entretanto, no segundo grupo, alguns participantes disseram que falar no telefone em público é menos ruim que usar um assistente, pois é possível camuflar o assunto da conversa falando apenas "sim" e "não", de forma a se sentirem menos expostos. Também foi considerado invasivo **ser obrigado a ouvir o que as outras pessoas estão pedindo no celular.**

Da mesma forma, o segundo grupo discutiu que o celular é o dispositivo mais pessoal que se têm, pois todos os seus dados estão ali. Por isso, não querem que outras pessoas ouçam ou tenham acesso ao que existe no aparelho. Ao se tratar da privacidade dos dados, para a maioria do segundo grupo, **existe mais receio das pessoas saberem as informações que estão no celular, que uma máquina saber o que tem ali.**

Ainda assim, a preocupação com a privacidade dos dados existe para alguns. Esses presam por uma maior transparência no que é coletado e no motivo pelo qual isso é feito, eles querem mais discussão sobre o assunto para estabelecer uma legislação que passe segurança; presam por saber quais dados serão coletados, como serão armazenados, distribuídos e, principalmente, quem vai ter acesso. Enquanto isso, outros se mostraram despreocupados com a questão; dizem já ter aceitado que seus dados estão sendo captados e não acreditam que seja de todo ruim, pois torna os sistemas mais personalizados e trazem algumas comodidades.

Apesar das duas visões, ambos acreditam que a privacidade dos dados já é um problema hoje. Por conta disso, os assistentes não estariam criando uma nova questão, mas sim agravando. Para a maioria, a transparência dos assistentes e o controle dos usuários são pontos essenciais. O que os assistentes coletam e inferem sobre seus usuários, bem como para o que será o usado e com quem essas informações estão sendo compartilhadas, devem poder ser acessados e alterados. Os participantes demonstraram que **querem ter controle sobre o assistente, e não vice-versa.**

Também foi levantado o ponto de se estar sendo escutado a todo momento. Mesmo quem disse acreditar que seus dados não são tão importantes caso vazem, admite que, em larga escala, existem pessoas que podem ser prejudicadas, como figuras públicas. Além disso, existe atenção com as empresas com quem os dados são compartilhados e **como isso será usado comercialmente**.

O receio é que, com os assistentes automatizando e tomando muitas decisões pelo usuário, possa existir um viés para anunciantes. Um exemplo dado foi o de pedir ao assistente para que compre um tênis: nesse caso, acreditam que o sistema irá considerar preferências específicas do usuário, mas também irá tentar encaixar o tênis da empresa com a qual tem parceria e pagou por um espaço publicitário.

Com relação a isso, os grupos também se dividiram. Alguns se sentem invadidos e se incomodam com os anúncios que recebem após ter falado algo próximo ao celular. Outros, já se sentem conformados com isso e acreditam que, se for para receber um anúncio, que pelo menos seja de algo que interessa.

Os mais preocupados reclamaram que, mesmo lendo e sabendo sobre a captação de seus dados nos termos e condições do serviço, e mesmo não gostando do que é solicitado, eles precisarão aceitar. "Ou você aceita ou se exclui de tudo e todos por não usar, e é da natureza humana querer se incluir".

Os menos preocupados acham que, apesar de ser necessária a cautela com a privacidade, acreditam que os assistentes irão facilitar a vida e "são legais", por isso deixam passar. Para eles, se fossem mais úteis, teriam ainda menos problemas com isso. Além disso, acreditam que, para se tornarem realmente personalizados, aprenderem com a forma do usuário falar e a forma de se relacionar com outras pessoas, haverá uma perda de privacidade; porém, acreditam que seja isso que fará a diferença para os assistentes ficarem bons.

Um dos participantes do segundo grupo retrucou que se sente mais incomodado com a perda de privacidade para os assistentes que para outras

ferramentas, como pesquisa, *Whatsapp* e outros, pois são ferramentas que considera necessárias. Com os assistentes virtuais, no entanto, se fosse preservar sua privacidade, preferia não usar, pois acredita que não fará falta. Mesmo que seja interessante, é sua privacidade que está em jogo.

A isso, outro participante rebateu que a maior preocupação é com quem tem acesso à informação e aos dados: se pertence a uma empresa, ao governo, ou se é algo que pode ser consumido por seres humanos, é perigoso. No entanto, se é algo que fica dentro de uma máquina, criptografado ou com sistema de segurança, é mais confiável.

Um ponto que foi levantado nos dois grupos e gerou discussão foi sobre o **compartilhamento desses dados com a polícia e o Estado**, e sobre até onde os assistentes devem dar-lhes acesso para tentar manter a segurança. O debate foi sobre a responsabilidade do assistente ao realizar uma busca ou tarefa para o usuário.

Quanto ao caso de um usuário fazer um pedido que é considerado ilegal ou potencialmente perigoso, como formas de cometer suicídio, os grupos se dividiram em dois posicionamentos. Enquanto uma parte dos participantes acreditam que deveria fazer parte de uma ética dos assistentes tratar esses casos, e não permitir ou incentivar nada que fosse ilegal ou errado; outros acreditam que não é uma incumbência dos assistentes tratar disso. Para esses, da mesma forma que as ferramentas de busca não julgam pirataria, por exemplo, a responsabilidade dos atos está no usuário.

No primeiro grupo, essa visão foi rebatida com o argumento que, hoje, o *Google* já trata alguns desses casos. "Para buscas sobre suicídio, por exemplo, ele entrega no resultado um número de telefone e *site* para que você possa procurar ajuda".

Outra preocupação que foi levantada em ambos os grupos foi sobre a personalização criar assistentes que **fomentem ideias ruins e reforcem**

comportamentos negativos. Não é desejável que uma pessoa com atitude racista molde um assistente que favoreça posturas antiéticas.

No segundo grupo, um participante também levantou que existe o risco não apenas na personalização, mas também em sistemas classificatórios. Dessa forma, o assistente poderia identificar pessoas de forma negativa através de dedução, além de favorecer preconceitos e estereótipos.

A desigualdade também foi foco de discussão no segundo grupo. O potencial alto custo desse tipo de tecnologia foi visto como uma possível forma de impedir o acesso de uma parte da população, ou permitir acesso apenas a uma versão da tecnologia menos avançada ou de baixa qualidade. Alguns participantes discordaram, acreditando que novas tecnologias vão diminuindo de custo e se tornando mais populares com o tempo. Apesar disso, outros participantes retrucaram que, historicamente, não é isso que ocorre, e ainda hoje existem pessoas sem acesso a *internet* no Brasil.

No segundo grupo, também houve o debate acerca de, no futuro, os assistentes tirarem cada vez mais o ser humano de suas funções, gerando **desemprego e desigualdade**. Por exemplo, se houvesse um assistente que funcionasse como professor, ele não só teria acesso a tudo sobre um conteúdo e sobre didática, mas também teria acesso às informações sobre o aluno, de forma a saber ensinar do melhor jeito para aquela pessoa. Alguns participantes viram isso como positivo, outros como negativo.

Quem viu de maneira negativa acredita que o fator humano é importante, que apenas um ser humano vai ter passado pelo que outro passou e vai ter empatia. Saberá não só que o raciocínio está errado, mas irá entender o motivo por trás disso. Por isso, ter alguém ensinando ainda será melhor.

Um participante disse acreditar que essa substituição irá gerar ainda mais desigualdade, pois afetará apenas serviços comuns. As profissões muito básicas

ou que desempenham serviços muito intelectualmente avançados não serão capazes ou financeiramente viáveis de serem substituídas.

Já os participantes que viram de forma positiva, acreditam que os assistentes serão mais capazes de desempenhar algumas funções e que isso não é um problema. As pessoas ganharão mais tempo com a evolução da tecnologia.

Em ambos os grupos, os participantes demonstraram desconfiança que essa nova tecnologia irá deixar as pessoas **mal acostumadas e preguiçosas**, tornando-as despreparadas para fazer as coisas sozinhas. Usaram como exemplo o uso dos celulares, onde a digitação fez com que as pessoas escrevessem errado e perdessem a prática de escrever a mão. Um dos participantes do segundo grupo, no entanto, disse que isso não é necessariamente ruim, pois atualmente as pessoas já dependem de muitas coisas.

Ainda no segundo grupo, os participantes se dividiram entre quem acreditava que esse tipo de tecnologia poderia levar a um **isolamento dos seres humanos** e quem achava que não. Quem acreditava no isolamento, defendia que isso irá acontecer pela falta da necessidade de interação social. "Está se criando uma tecnologia que se parece com uma pessoa para que não exista a necessidade de falar com uma pessoa de verdade". Eles se preocupam que, em um futuro mais distante, as pessoas acabem criando relações tão fortes, ou mais fortes, com um assistente virtual que com uma pessoa.

Quem discordava, acredita que a tecnologia irá aliviar tarefas trazendo mais tempo para as interações humanas que realmente importam. Atualmente, existe uma resistência em precisar falar e interagir com as pessoas, pois se usa muita mensagem, e talvez os assistentes virtuais sejam uma forma de voltar a naturalizar a fala.

Apesar disso, nos dois grupos, a maioria dos participantes acredita que os assistentes não tem uma missão única de execução de tarefas, mas também têm uma **função social**. Nos dois grupos usaram o exemplo do Japão, pontuando que,

onde há muitos idosos, eles estão ficando sozinhos e doentes, e acabam se suicidando porque não tem com quem conversar e conviver. A função social e de convivência dos assistentes é, por tanto, importante para quem não tem a possibilidade de sociabilização.

Os grupos mostraram também que existem pessoas que são mais introvertidas e não querem ter alguém durante toda a vida as acompanhando, pois preferem ter tempo em casa para ficarem sozinhas. No entanto, na percepção delas, quando algo "cai no uso", vira uma tendência obrigatória. Para essas pessoas mais introvertidas, existe uma preocupação de **serem obrigadas a terem um assistente dentro de sua casa porque todos os móveis e toda a sociedade estão conectados dessa forma.**

Em ambos os grupos, a maioria chegou à conclusão que os assistentes virtuais têm potencial para serem ferramentas incríveis, e que irão evoluir e estarão cada vez mais presentes no nosso cotidiano. A maioria acredita que irá usá-los no futuro, mas não acham que possuem tanto valor hoje, sendo mais divertidos que úteis.

O participante que mais demonstrou preocupação com o uso dos assistentes estava no segundo grupo e, ao final da discussão, disse saber que as coisas, quando muito novas, dão medo no início. Ele afirmou que talvez venha a usá-los no futuro: "Podem haver coisas que, hoje, eu ache que não vou gostar, mas que no futuro, quando acontecer, eu posso ser totalmente a favor".

Ainda assim, uma minoria dos grupos acredita que existam mais pontos negativos que positivos e que, infelizmente, achar bom ou ruim já não é uma questão, pois vai acontecer. Apenas um participante de cada grupo disse que acredita que não usará, mesmo mais para frente, essa tecnologia.

De todas as maneiras, foi ressaltado que é uma tecnologia nova e que ainda existem muitos pontos que precisam ser discutidos, como a relação com os problemas sociais, a ética e a privacidade.

5

Considerações finais e desdobramentos

A premissa dessa pesquisa tratou da personificação e do uso de interfaces conversacionais nos assistentes virtuais, acreditando que esses traços facilitariam a sua aceitação e tornariam a opinião sobre os mesmos positiva. Ao final do estudo, o resultado demonstrou que a personalidade desses assistentes realmente facilita a sua adoção e tem efeitos positivos sobre a percepção do usuário, porém as interfaces de voz ainda geram desconforto quando usadas em ambiente público. As maiores preocupações levantadas pelos usuários em relação aos assistentes de voz não foram técnicas ou de usabilidade, mas reflexões sobre o impacto dos assistentes na sociedade e na vida das pessoas.

Outras questões norteadoras e objetivos que foram levantados no planejamento da pesquisa são aqui respondidos a partir dos resultados obtidos:

- **Qual o impacto da fala na experiência de uso de Assistentes Virtuais Personificados? O uso da voz, no lugar de texto escrito como acontece em chatbots, é entendido como benéfico pelo seu usuário?**

Apesar da fala como interface ter benefícios percebidos, como a possibilidade de interação enquanto as mãos estão ocupadas, como o encurtamento de caminhos de navegação da interface e como a possibilidade de melhoria para a acessibilidade; os usuários ainda se sentem muito incomodados ao usar esses assistentes em lugares públicos e ao interagir com objetos inanimados (computadores e celulares), o que pode gerar uma dificuldade na sua adoção.

- **Os estudos levantados durante a revisão bibliográfica refletem as particularidades e casos de uso do público estudado? Quais as diferenças e pontos de atenção quando se trata desse público?**

Os estudos levantados refletem, em grande parte, os casos de uso do público estudado, principalmente os pontos retratados por Nass e Brave (2007). No entanto, existem três particularidades que podem ser observadas: a preocupação com a segurança ao usar celular em público, a atenção às questões sociais e éticas, e a preferência por vozes femininas. Esses pontos ou não foram mencionados ou discordaram da bibliografia estudada, e são possíveis desdobramentos dessa pesquisa.

- **Esse público usa ou pretende usar essa interface no futuro? Como está se dando adoção e quais são as opiniões acerca dos Assistentes Virtuais Personificados?**

A maioria dos participantes que usa atualmente os assistentes, o fazem por razões de preguiça e entretenimento, sendo os AVPs analisados como pouco úteis hoje. Quanto ao futuro, a grande maioria acredita no potencial dessa tecnologia, e considera sua adoção e uso conforme os assistentes forem evoluindo e se tornando mais comuns. O cenário conectado se mostra como um grande influenciador nesse último caso.

Tanto nos resultados do questionário, quanto nos resultados dos grupos de foco, foi fácil perceber que a personificação e o uso da interface conversacional influenciam de maneira positiva na adoção e na percepção de experiência. Os AVPs, assistentes que são personificados e usam interfaces conversacionais, tinham uma maior adoção e eram mais bem avaliados.

No entanto, foi possível notar que esses traços se influenciam. Os assistentes que só adotavam uma persona, sem o uso da interface conversacional, como é o caso dos assistentes virtuais de *telemarketing*, acabam por gerar uma quebra de expectativa e uma frustração, que culminam em uma percepção negativa dos mesmos.

Os dois AVPs que foram o foco dessa pesquisa, a Google Assistente e a Siri, foram avaliadas de maneiras diferentes pelos usuários. A Siri teve sua avaliação prejudicada por se considerada mais limitada, por ter problemas de

usabilidade e falhas na sua interface conversacional. Quanto a usabilidade, a falta de *feedback* quando há indisponibilidade de *internet* foi seu maior problema. Na aplicação da interface conversacional, a necessidade de invocação para toda nova interação e a falta de continuidade dos diálogos, por conta de não resgatar informações sobre os enunciados anteriores, foram pontos críticos. Já a Google Assistente foi avaliada de maneira positiva, pois não apresentava esses problemas.

Conquanto, houveram percepções divididas sobre a personalidade dos dois AVPs, que parecem se relacionar com a identificação social dos usuários. Ou seja, usuários mais extrovertidos parecem qualificar de maneira mais positiva a Siri e menos positiva a Google Assistente, enquanto usuários mais introvertidos parecem fazer o oposto. Esse é um tópico interessante que poderia ser melhor explorado em um desdobramento da pesquisa.

É importante pontuar que essa percepção negativa da Siri pode ter sido intensificada porque a maioria dos participantes usava a Siri na sua versão em português brasileiro, que é mais limitada que sua versão original em inglês. Apesar disso, a comparação continua se mostrando verdadeira, pois o uso da Google Assistente também ocorreu na versão em português, mais restrita que a sua versão original, também em inglês.

Além disso, é relevante destacar que ainda há uma grande discussão que precisa ser aprofundada quanto aos aspectos éticos e sociais dos AVPs. Esse é, também, um desdobramento interessante para a pesquisa.

Referências bibliográficas

2001: A Space Odyssey. Direção de Stanley Kubrick. [s.l.] Metro-Goldwyn-Mayer (MGM), 1968.

ADAMS, D. **O Guia do Mochileiro das Galáxias.** 1 ed. [s.l.] Arqueiro, 2009.

ASIMOV, I.; PEREIRA, A. S. **Eu, Robô.** 1 ed. [s.l.] Aleph, 2014.

EX Machina. Direção de Alex Garland. [s.l.] Universal Pictures, 2015.

GALEGO, Carla; GOMES, Alberto. **Emancipação, ruptura e inovação: o “focus group” como instrumento de investigação.** Revista Lusófona de Educação, [S.l.], v. 5, n. 5, nov. 2005. ISSN 1646-401X. p. 173-184
Disponível em:
<<http://revistas.ulusofona.pt/index.php/rleducacao/article/view/1012>>.
Acesso em: 20 mai. 2018.

GOMES, R. J. R. **Teste de interfaces de voz.** Porto: Universidade do Porto, maio 2007.

GREENFIELD, A. **Everyware: The Dawning Age of Ubiquitous Computing.** 1. ed. Califórnia: New Riders Publishing, 2006.

GRICE, H. P. Logic and Conversation. In: COLE, P.; MORGAN, J. L. **Syntax and Semantics 3: Speech Acts.** Nova Iorque: Academic Press, 1975. p. 45-58.

HER. Direção de Spike Jonze. [s.l.] Annapurna Pictures, 2013.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Acesso à Internet e à televisão e posse de telefone móvel celular para uso pessoal 2017.** Disponível em:
<https://biblioteca.ibge.gov.br/visualizacao/livros/liv101631_informativo.pdf>.
>. Acesso em: 10 jan. 2019.

IRON Man. Direção de Jon Favreau. [s.l.] Paramount Pictures, 2008.

MCTEAR, M.; CALLEJAS, Z.; BARRES, D. G. **The Conversational Interface: Talking to Smart Devices.** 1 ed. Nova Iorque: Springer, 2016.

MOGGRIDGE, B. **Designing Interactions**. 1 ed. Cambridge, Massachusetts: The MIT Press, 2007.

NASS, C.; BRAVE, S. **Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship**. Cambridge, Massachusetts; Londres: The MIT Press, 2007. Paginação irregular.

NORMAN, D. A. **Design Emocional: Por que adoramos (ou detestamos) os objetos do dia-a-dia**. Tradução Ana Deiró. Rio de Janeiro: Rocco, 2008.

NORMAN, D. A. **Emotional Design: Why We Love (or Hate) Everyday Things**. Nova Iorque: Basic Books, 2005.

NORMAN, D. A. **Design Emocional: Por que adoramos (ou detestamos) os objetos do dia-a-dia**. Tradução Ana Deiró. Rio de Janeiro: Rocco, 2008.

PASSENGERS. Direção de Morten Tyldum. [s.l.] Columbia Pictures Corporation, 2016

PEARL, C. **Designing Voice User Interfaces: Principles of Conversational Experiences**. 1 ed. Beijing: O'Reilly Media, 2016. Paginação irregular.

SAFFER, D. **Designing for Interaction: Creating Innovative Applications and Devices**. 2 ed. Califórnia: New Riders Publishing, 2009.

THE SPACE Between Us. Direção de Peter Chelson. [s.l.] STX Entertainment, 2017.

WEISER, M. The Computer for the 21st Century. **Scientific American**, v. 265, n. 3, p. 94-104, 1 set. 1991.

WEISER, M. Hot topics - Ubiquitous Computing. **Computer**, v. 26, n. 10, p. 71-72, out. 1993.

WEISER, M. The World is Not a Desktop. **Interactions**, v. 1, n. 1, p. 7-8, jan. 1994.

WEISER, M.; BROWN, J. S. Designing Calm Technology. **Powergrid Journal**, v. 1, 1996.

WEISER, M.; BROWN, J. S. The coming age of calm technology. In: DENNING, P. J.; METCALFE, R. M. (Eds.) . **Beyond Calculation**. Nova Iorque: Copernicus, 1997. p. 75-85.

7 Apêndices

7.1. Questionário online

formulário inicial

Assistentes virtuais e a voz como interação

O Sr. (a) está sendo convidado (a) como voluntário (a) a participar da pesquisa **“Assistentes virtuais personificados e a voz como forma de interação”**. O objetivo da pesquisa é obter informações junto aos usuários de assistentes virtuais e interfaces de voz no Brasil sobre a adoção e usos dos mesmos, bem como trazer à discussão os problemas e oportunidades encontrados em seu uso hoje, e intenções futuras de uso.

Esta é uma pesquisa de Mestrado, realizada pela mestranda Marcela Pedroso Maués e orientada pela Profa. Cláudia Mont’Alvão, D.Sc. Qualquer dúvida a respeito dos procedimentos, dos resultados e/ou de assuntos relacionados à pesquisa será esclarecida pela orientadora Cláudia Mont’Alvão, no telefone (21) 3527-1594 ou através dos emails: cmontalvao@puc-rio.br, ergonomiapuc@gmail.com.

O pesquisador irá tratar a sua identidade com sigilo, e serão adotados os procedimentos para resguardar a identidade dos entrevistados.

Dados do participante

Onde você mora atualmente? (estado)

Qual o seu gênero?

- ☐ Homem
- ☐ Mulher
- ☐ Prefiro não responder
- ☐ Outro: _____

Qual o seu grau de escolaridade?

- ☐ Ensino fundamental / ginásio incompleto
- ☐ Ensino fundamental completo
- ☐ Ensino médio / 2º grau incompleto
- ☐ Ensino médio completo
- ☐ Ensino superior incompleto
- ☐ Ensino superior completo
- ☐ Pós-graduação ou especialização incompleto
- ☐ Pós-graduação ou especialização completo

Qual a sua idade?

Sobre o uso de assistentes virtuais

Através de que meios você acessa a internet? (marque quantas desejar)

- ☐ Celular
- ☐ Tablet
- ☐ Computador desktop
- ☐ Laptop
- ☐ GPS
- ☐ Videogame
- ☐ Smartwatch
- ☐ SmartV
- ☐ ChromeCast

- ☐ Amazon Echo
- ☐ Google Home
- ☐ HomePod
- ☐ Outro: _____

Nos últimos 3 meses, você usou algum assistente virtual? Para responder essa pergunta, considere "assistente virtual" qualquer sistema que mimetize a interação com uma pessoa, como assistentes virtuais de telemarketing, Siri, Cortana, Google Assistant, Bixby, Alexa, assistentes virtuais que respondem via chat (chatbots) etc.

- ☐ Sim
- ☐ Não
- ☐ Não sei responder

Qual assistente virtual você usou nos últimos 3 meses? (marque quantas desejar)

- ☐ Assistente virtual de telemarketing/telefonia
- ☐ Chatbot (assistente virtual que responde via chat) em um website ou aplicativo
- ☐ Siri (assistente da Apple/iOS)
- ☐ Google Assistant (assistente da Google/android)
- ☐ Bixby (assistente da Samsung)
- ☐ Cortana (assistente da Microsoft/ Windows)
- ☐ Alexa (assistente da Amazon)
- ☐ Outro: _____

Nos últimos 3 meses, em qual dispositivo você usou o assistente virtual? (marque quantas desejar)

- ☐ Usei no celular
- ☐ Usei no tablet
- ☐ Usei no computador desktop
- ☐ Usei no laptop
- ☐ Usei no gps
- ☐ Usei no videogame
- ☐ Usei no smartwatch

- ☐ Usei na smarTv
- ☐ Usei no Chromecast
- ☐ Usei no Amazon Echo
- ☐ Usei no Google Home
- ☐ Usei no HomePod
- ☐ Outro: _____

Com que fins você usou/usa o assistente virtual? (marque quantas desejar)

- ☐ Para saber as horas
- ☐ Para saber sobre a previsão do tempo
- ☐ Para saber sobre as últimas notícias
- ☐ Para saber sobre o trânsito
- ☐ Para fazer uma pergunta/pesquisa
- ☐ Para ligar para alguém
- ☐ Para enviar uma mensagem para alguém
- ☐ Para salvar um compromisso na agenda
- ☐ Para armar um cronômetro ou despertador
- ☐ Para iniciar um jogo no console
- ☐ Para buscar uma rota para um lugar
- ☐ Para colocar uma música para tocar
- ☐ Para colocar um filme/série
- ☐ Para controlar um filme/série que estava tocando
- ☐ Para controlar o ambiente (por exemplo acender uma luz, ligar o ar condicionado, trancar uma porta)
- ☐ Para me auxiliar enquanto cozinhava
- ☐ Para solicitar ajuda em um website ou aplicativo
- ☐ Por curiosidade, para explorar o assistente
- ☐ Outro: _____

Como foi sua experiência com o assistente?

- ☐ Foi boa, ainda faço uso.
- ☐ Foi boa, mas não continuo usando.
- ☐ Indiferente, mas ainda faço uso.
- ☐ Indiferente, não continuo usando.
- ☐ Foi ruim, mas ainda uso.
- ☐ Foi ruim, não continuo usando.

Você gostaria de deixar algum comentário referente à pesquisa ou ao formulário?

Você gostaria de ajudar a pesquisa sobre assistentes virtuais participando de um grupo de foco?

- ☐ Sim
☐ Não

Informações para recrutamento

Deixe aqui suas informações de contato para que você possa ser chamado para participar do grupo de foco.

E-mail para contato:

Telefone para contato:



TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

O Sr. (a) está sendo convidado (a) como voluntário (a) a participar da pesquisa **“Assistentes virtuais personalizados e a voz como forma de interação”**. Esta é uma pesquisa de Mestrado, realizada pela mestrandia Marcela Pedrosa Maués e orientada pela Profa. Cláudia MontAlvão, D.Sc. Qualquer dúvida a respeito dos procedimentos, dos resultados e/ou de assuntos relacionados à pesquisa será esclarecida pela orientadora Cláudia MontAlvão, no telefone (21) 3527-1594 ou através dos e-mails: montalvao@puc-rio.br, ergonomiapuc@gmail.com. O objetivo da pesquisa é obter informações junto aos usuários de assistentes virtuais e interfaces de voz sobre a adoção e usos dos mesmos, bem como trazer à discussão os problemas e oportunidades encontrados em seu uso hoje, e intenções futuras de uso. O pesquisador irá tratar a sua identidade com sigilo, e serão adotados os procedimentos para resguardar a identidade dos entrevistados.

Para este estudo adotaremos os seguintes procedimentos em **2 (duas) etapas** e você participará das duas, sendo a segunda etapa **pessoalmente**, em contato direto com o pesquisador:

QUESTIONÁRIO: você será convidado a responder sobre seu uso, sua opinião e suas percepções acerca de um assistente virtual que usa a voz como forma de interação. Nessa etapa, você não entrará em contato com outros participantes, e o pesquisador ou assistente de pesquisa irá anotar seus comentários.

GRUPO DE FOCO: após a entrevista, em um dia a combinar, você se juntará em um grupo de 6 a 8 pessoas para debater sobre os usos, opiniões e percepções atuais sobre um assistente virtual que usa interface de voz. Essa etapa será registrada por meio de vídeo e áudio para estudo e análise. Nenhuma imagem, vídeo ou áudio será usado fora dessa pesquisa.

BENEFÍCIOS: Ao final da pesquisa pretende-se estabelecer as percepções e opiniões do brasileiro residente do Rio de Janeiro acerca de assistentes virtuais que usam a voz como interface, bem como as razões que o levam a gostar ou desgostar da interação, usos atuais e intenções de uso futuras. Com isso, pretende-se ajudar futuros projetistas a desenhar melhores interações com assistentes virtuais e com interfaces de voz.

RISCOS: por se tratar de uma metodologia cujas técnicas pautam-se em levantamento de dados em questionário e percepções sobre um sistema, a mesma é considerada com possibilidade quase nula de riscos para o participantes. Serão adotados os procedimentos para resguardar a identidade dos participantes, para que não haja algum constrangimento pelo fato de alguém reconhecer sua fala no decurso da pesquisa ou em sua publicação.

RESSARCIMENTO: Os voluntários dessa pesquisa não serão remunerados por essa participação e nem tampouco correrão riscos.



Para participar deste estudo você não terá nenhum custo, nem receberá qualquer vantagem financeira. Você será esclarecido sobre o estudo em qualquer aspecto que desejar e estará livre para participar ou recusar-se a participar. Poderá retirar seu consentimento ou interromper a participação a qualquer momento. A sua participação é voluntária e a recusa em participar não acarretará qualquer penalidade ou modificação na forma em que é atendido pelo pesquisador.

O pesquisador irá tratar a sua identidade com padrões profissionais de sigilo.

Os resultados da pesquisa estarão à sua disposição quando finalizada. Seu nome ou o material que indique sua participação não será liberado sem a sua permissão.

O (A) Sr. (a) não será identificado em nenhuma publicação que possa resultar deste estudo.

Este termo de consentimento encontra-se impresso em duas vias, sendo que uma cópia será arquivada pelo pesquisador responsável, e a outra será fornecida a você.

Eu, _____ portador do documento de identidade _____ fui informado (a) dos objetivos do estudo “Assistentes virtuais personalizados e a voz como forma de interação” , de maneira clara e detalhada e esclareci minhas dúvidas. Sei que a qualquer momento poderei solicitar novas informações e modificar minha decisão de participar se assim o desejar.	_____
Declaro que concordo em participar desse estudo. Recebi uma cópia deste termo de consentimento livre e esclarecido e me foi dada a oportunidade de ler e esclarecer as minhas dúvidas.	Rio de Janeiro, _____ de _____ de 2018.
_____	Assinatura voluntário
_____	Assinatura pesquisador

7.3. Roteiro do grupo de foco

Roteiro do Grupo de Foco

Assistentes virtuais e a voz como interação

Introdução

sobre o estudo e assinatura do termo de consentimento

Bom dia / Boa tarde, meu nome é Marcela e eu irei moderar o grupo de foco sobre assistentes virtuais hoje. O grupo de foco é uma técnica de pesquisa onde reunimos pessoas para discutir um tópico, que hoje é assistentes virtuais. A ideia é entender as diferentes opiniões e colocá-las em debate para aprofundar o conhecimento sobre as diferenças e semelhanças de cada visão.

É muito importante todos entenderem que não existe um lado certo ou um errado, e que ninguém aqui está sendo testado, só os assistentes. Também é importante destacar que as sessões serão gravadas por áudio e vídeo para melhor analisar depois, mas nenhuma imagem ou áudio vai ser usado para outra finalidade que não esse estudo de mestrado. Dessa forma, para preservar a privacidade de vocês, todas as identidades, nomes, imagens e gravações de áudio serão ocultados e tratados como anônimos. As impressões gerais serão tratadas como resultado apenas, sem citar ninguém.

Para dar essa segurança a vocês, eu irei pedir para que vocês preencham esse termo de consentimento de participação na pesquisa. Vocês podem ler à vontade e uma via ficará com vocês. Qualquer dúvida, estarei a disposição para responder.

Primeira parte sobre os assistentes virtuais hoje

1. Primeiro, para ter certeza que todos iremos partir do mesmo ponto para a discussão, iremos testar algumas perguntas para a Assistente Virtual do iOs e para a Assistente do Google.

Perguntar ao assistente virtual:

- . Quem é você?
- . Quais livros Stephen King escreveu?
- . Como está a previsão do tempo para hoje?
- . Ligue para minha mãe
- . Mande uma mensagem para meu namorado
- . Como está o trânsito até o trabalho?
- . Como faço para ir até a PUC-Rio?

O que vocês acharam dessas perguntas? Já tinham usado alguma dessas funções? Aconteceu o que vocês esperavam?

2. Qual (ou quais) assistente virtual vocês usaram recentemente? Como foi a experiência de vocês com esse(s) assistente(s), vocês gostaram? Por que vocês gostaram ou não?
 - a. **Se alguém usou mais de um:** qual você gostou mais? Por quê? Você percebe alguma semelhança ou diferença entre eles?
3. Em que língua estava o assistente que vocês usaram? Onde vocês usaram esse assistente (por qual meio)?

4. Com quais finalidades vocês usaram o assistente?
 - a. O que vocês esperavam que o assistente fizesse? E o que ele fez? (*observar expectativas x realidade*)
5. Quais os problemas vocês vêm nesses assistentes de hoje?
 - a. E com relação à privacidade, isso preocupa vocês?
6. Vocês lembram como era a voz desse assistente? Era masculina ou feminina?
 - a. Por que vocês acham que era assim?
 - b. Vocês gostaram da voz do assistente?
7. E com relação à personalidade, você lembra como era a do assistente que vocês usaram? Se era simpático, debochado, mais formal, amigável...?
 - a. Por que vocês acham que era assim?
 - b. O que vocês acharam disso? Vocês gostaram?
8. Vocês continuam usando esse assistente?
 - a. Se sim, com qual frequência?
 - b. Se não, por quê?

Segunda parte sobre expectativas futuras sobre os assistentes virtuais

1. Agora, nós iremos falar um pouco sobre o futuro desses assistentes e expectativas para eles. Alguém aqui já viu o filme Her? E a última apresentação do Google sobre o Google Duplex, com aquele assistente que marca um horário no salão e no restaurante?

Mostrar vídeos:

- . Cena de instalação da Samantha do filme Her
- . Vídeo de apresentação do Google Duplex durante a apresentação do Google IO de 2018

Caso todos já tenham visto o vídeo, não é necessário mostrar novamente.

O que vocês acharam dessas visões de futuro?

- a. Vocês gostaram de algum desses assistentes? Por quê?
 - b. Com que finalidade vocês usariam esses assistentes? Para que vocês acham que eles seriam úteis?
2. Como vocês acham que serão esses assistentes no futuro? O que vocês vislumbram de possibilidades para eles?
 - a. O que vocês gostariam que eles fizessem?
 - b. E o que vocês gostaria, que eles não fizessem?
3. Que problemas vocês acham que esses assistentes podem encontrar no futuro? Que problemas que esses assistentes podem vir a ter?
 - a. Na sua opinião, esses problemas podem impedir a adoção desses assistentes?

4. Como vocês acham que deveria ser a voz desses assistentes no futuro? E a personalidade?
 - a. Por que vocês acham que deveria ser assim?
 - b. Deveria ser assim sempre, em todas as situações? Quando não deveria ser?
5. Vocês gostariam de adicionar mais alguma informação ou fazer algum outro comentário? Vocês gostaram da discussão?

Encerramento

Eu agradeço muito a participação de vocês. Vocês gostaram da discussão? Espero que tenha sido proveitosa para vocês como foi pra mim. Qualquer dúvida ou outro ponto que vocês lembrem mais tarde e queiram acrescentar, esse é o meu e-mail e telefone de contato. Muito obrigada.