



Lucas José da Silva Santos

**Interfaces conversacionais:
A experiência no uso da voz em interações**

DISSERTAÇÃO DE MESTRADO

Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre em *Design* ao Programa de Pós-graduação em *Design* da PUC-RIO.

Orientadora: Prof. Dra. Maria Manuela Rupp Quaresma

Rio de Janeiro, 16 de setembro de 2020.



Lucas José da Silva Santos

**Interfaces de voz:
A experiência no uso da voz em interações**

Dissertação apresentada ao Programa de Pós-graduação em *Design* da PUC-RIO como parte dos requisitos para a obtenção do título de Mestre em *Design*.

Profa. Maria Manuela Rupp Quaresma
Orientadora

Departamento de Arte & *Design* - PUC-Rio

Prof. Marcelo Fernandes Pereira
Departamento de Arte & *Design* - PUC-Rio

Profa. Barbara Emanuel
Departamento de Comunicação Social - UFF



Lucas José da Silva Santos

Interfaces de voz:

A experiência no uso da voz em interações

Dissertação apresentada ao Programa de Pós-graduação em *Design* da PUC-RIO como parte dos requisitos para a obtenção do título de Mestre em *Design*.

Profa. Maria Manuela Rupp Quaresma

Orientadora

Departamento de Arte & *Design* - PUC-Rio

Rio de Janeiro, 16 de setembro de 2020.

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

Lucas José da Silva Santos

Formou-se em Desenho Industrial, com habilitação em Mídias Digitais, pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) em 2015. Tem experiência profissional desde 2014, atuando como *Product Designer* em projetos centrados nos usuários dentro da área de Tecnologia, desenvolvendo produtos e serviços digitais. Atualmente dedica-se ao mestrado e ao mercado de trabalho, sendo nesse último líder de um time de *design*.

Ficha Catalográfica

Santos, Lucas

Interfaces conversacionais : a experiência no uso da voz em interações / Lucas José da Silva Santos ; orientadora: Maria Manuela Rupp Quaresma. – 2020.

104 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Artes e Design, 2020.

Inclui bibliografia

1. Artes e Design – Teses. 2. Interação humano-máquina. 3. Experiência do usuário. 4. Teoria da conversação. 5. Interface conversacional. 6. Interface de voz. I. Quaresma, Maria Manuela

CDD: 700

Agradecimentos

Primeiramente, gostaria de agradecer a todas as pessoas que integram a comunidade LGBTQIA+ da qual faço parte, por resistirem a tantos anos de opressão e marginalização da sociedade que, por muitas vezes, nos impedem de alcançar espaços como o da academia. Agradeço especialmente a todas as travestis pretas que pertencem à linha de frente pela nossa sobrevivência. A vocês, muito obrigado por cada luta, lágrima e suor derramados para que pessoas da nossa comunidade pudessem alcançar objetivos cada vez maiores.

À minha mãe, Lucia Aparecida, mulher preta que nunca hesitou em me apoiar ao longo dos meus 27 anos. Que sempre me fortaleceu e que é e sempre será minha maior fonte de inspiração e força.

À família incrível de mulheres pretas à qual pertenço, por me mostrar que a academia também é um espaço para nós, pretos, por mais que as estatísticas nos digam o contrário.

Ao Davi de Castro, que durante mais da metade destes dois anos, foi o amor da minha vida e me estimulou e apoiou com graça, carinho e afínco.

À Rafaela Nobrega, minha grande amiga, que me apoiou em cada etapa desta pesquisa, com todo seu conhecimento e amor.

Às minhas irmãs, a Matheus Valois, a Gabriel Faria e a Rodrigo Moreira, que foram por muitas vezes meu colo e porto seguro nessa jornada difícil.

Ao CNPQ e à PUC-Rio pelos auxílios concedidos, que possibilitaram a realização desta pesquisa.

À minha orientadora, Manuela Quaresma, pela dedicação excepcional e colaboração em todo processo da minha pesquisa de mestrado.

Resumo

Santos, Lucas; Quaresma, Manuela (orientadora). **Interfaces conversacionais: a experiência no uso da voz em interações.** Rio de Janeiro, 2020. Dissertação de mestrado - Departamento de Artes e Design, Pontifícia Universidade Católica do Rio de Janeiro.

O presente estudo parte da premissa de que novas formas de se comunicar surgem ao longo dos anos, inclusive na relação entre o humano e a máquina, onde, atualmente, a fala é o veículo comunicacional predominante. E nesse diálogo há uma busca para a adequação da máquina à nossa forma de comunicação e uma constante necessidade de adequação dos seres humanos à forma de comunicação da máquina. Essa investigação visa elucidar se sistemas baseados em interfaces de voz compõe o modelo de interação fundado no que se define como conversa, sendo amparados pelas áreas da ergonomia e usabilidade e por fundamentos provenientes da teoria da conversação. Este trabalho encontra-se estruturado em quatro seções: a primeira equivale a fundamentação teórica; a segunda seção abrange o levantamento de dados através de entrevistas com usuários de interfaces conversacionais; a terceira seção constitui-se na análise de conteúdo das entrevistas; e por fim, a quarta seção envolve as considerações em torno das experiências no uso de interfaces conversacionais atuais.

Palavras-chave

Interação Humano-Máquina; Experiência do Usuário; Teoria da Conversação; Interface Conversacional; Interface de Voz; Interação Sonora; Assistente Virtual Inteligente.

Abstract

Santos, Lucas; Quaresma Manuela (adviser). **Conversational Interfaces: the experience in the use of voice interactions**. Rio de Janeiro, 2020. Masters Dissertation - Departamento de Artes e Design, Pontifícia Universidade Católica do Rio de Janeiro.

The present study starts from the premise that new ways of communication are emerging throughout the years, including the relationship between human and machine, which nowadays, speech is the predominant communicational vehicle. And in this dialogue, there is a search for the adaptation of the machine to our form of communication and a constant need to human's adaptation to the machine's form of communication. This investigation attempts to clarify if systems based on voice interfaces are in fact composed by a model of interaction that is defined as conversation, and if they are actually being supported by the areas of ergonomics, usability, and also by the basics fundamentals of conversation theory. This work finds itself structured in four sections: the first one corresponds to the theoretical foundations; the second section embraces data collection, which gathered several interviews with the users of conversational interfaces; the third section is established on the content analysis of the previous interviews; lastly, the fourth section concerns the final considerations of the current experience of conversational interfaces.

Key-words

Human-Machine Interaction; Users Experience; Conversation Theory; Conversational Interface; Voice Interface; Sound Interaction; Intelligent Virtual Assistant.

Sumário

1. Introdução	1
1.1. Delineamento do contexto	1
1.2. Objetivos, roteiro metodológico e justificativa da investigação	5
1.3. Estrutura da dissertação	7
2. Sociedade e comunicação	9
2.1. A comunidade primitiva	9
2.2. A sociedade tradicional	12
2.3. A sociedade complexa	14
2.4. A explosão da <i>internet</i> como meio comunicacional	16
3. <i>Design</i> de interação: sistemas conversacionais	19
3.1. O que é uma interface?	19
3.2. Do termo <i>design</i> de interação à definição de interação	21
3.2.1. O que é interação?	22
3.2.2. O papel das atividades mentais nas interações	29
3.2.2.1. Captação de informações	30
3.2.2.2. Memória	31
3.2.2.3. Atenção prolongada (vigilância)	32
4. Interação sonora	35
4.1. Cognição	35
4.2. Interação por voz	38
4.2.1. O poder das interações na segunda era das interfaces por voz	42
4.3. Conversa como elemento primordial da interação	44
5. Delineamento da pesquisa	50
5.1. Tema	50

5.2. Problema	50
5.3. Hipótese	52
5.4. Objeto de pesquisa	52
5.5. Objetivos	52
5.5.1. Objetivo geral	52
5.5.2. Objetivos específicos	53
5.6. Justificativa e relevância da pesquisa	53
5.7. Métodos, técnicas e procedimentos	54
5.7.1. Universo de entrevista (entrevista semiestruturada)	54
5.8. Análise de conteúdo como ferramenta analítica	57
5.8.1 Descrição do processo adotado	57
6. Análise de conteúdo das entrevistas	61
6.1. Unidades de registro, unidades de contexto e categorias extraídas das entrevistas	61
6.2. Análise de conteúdo e inferências	61
6.3. Motivo de uso de sistemas conversacionais	63
6.4. Necessidade de uso	64
6.5. Eficácia na realização de comandos	65
6.6. Experiência no uso de sistemas conversacionais	66
6.6.1. Pontos positivos	67
6.6.2. Pontos negativos	68
6.7. Repetição de comandos	69
6.8. Turnos	69
6.9. Compreensão da linguagem (fala) pelo sistema	70
6.10. Tempo de escuta pelo sistema	71
6.11. Experiência de conversa	72
6.12. Percepção de aprendizagem por parte do sistema	73
7. Discussão	75
7.1. Contexto de comando	75
7.2. Linguagem	77
7.3. De troca para comando	79
7.4. Acordo	79
7.5. De ação ou (trans)ação para resposta-ação	79

7.6. <i>Feedback</i> /Percepção	80
7.7. Revisão do modelo conversacional	81
8. Conclusão e desdobramentos futuros	83
8.1. Percurso percorrido ao longo da dissertação	84
8.2. Conclusão	85
8.3. Desdobramentos futuros	87
9. Referências bibliográficas	89
Apêndice 1	94

Lista de figuras

Figura 01 - Google Assistente	1
Figura 02 - <i>Siri</i>	1
Figura 03 - <i>Amazon Echo</i>	2
Figura 04 - <i>Google Home</i>	2
Figura 05 - Cena do Filme <i>Nós</i> (2019) - parte I	3
Figura 06 - Cena do Filme <i>Nós</i> (2019) - parte II	3
Figura 07 - Modelo de <i>Feedback</i>	24
Figura 08 - Modelo de Abismo	25
Figura 09 - Sete Estágios de Ação	26
Figura 10 - Modelo de como sente-conhece-faz	27
Figura 11 - Ergonomia humano-máquina	28
Figura 12 - Evolução das linguagens	40
Figura 13 - Modelo de Conversação de Paul Pangaro	46
Figura 14 - Revisão modelo conversacional	82

Lista de tabelas

Tabela 01 - Estrutura para análise de conteúdo	61
Tabela 02 - Motivo de uso de sistemas conversacionais	64
Tabela 03 - Necessidade de uso	64
Tabela 04 - Eficácia na realização de comandos	66
Tabela 05 - Experiências no uso de sistemas conversacionais	66
Tabela 06 - Pontos positivos no uso de sistemas conversacionais	67
Tabela 07 - Pontos negativos no uso de sistemas conversacionais	68
Tabela 08 - Repetição de Comando ao sistema para conclusão de tarefas	69
Tabela 09 - Quantidade de turnos por interação	70
Tabela 10 - Compreensão da Linguagem pelo Sistema	71
Tabela 11 - Tempo de escuta pelo sistema	72
Tabela 12 - Experiência de conversa	73
Tabela 13 - Aprendizado do sistema	74

Introdução

1.1

Delineamento do contexto

Atualmente, o universo de produtos baseados em interação por voz vive uma grande expansão. Com mais de 146,9 milhões de alto-falantes inteligentes vendidos em 2019 - representando um aumento de 70% se comparado ao ano de 2018 (Strategy Analytics, 2019), é possível observar que cada vez mais pessoas estão interagindo através da voz com sistemas e plataformas. Hoje, é possível solicitar um Uber por comando de voz através de um alto-falante inteligente ou de um assistente virtual. A Microsoft (2019) afirma, numa pesquisa realizada seis meses antes de sua divulgação, que 72% dos entrevistados relataram utilizar assistentes virtuais, comprovando a democratização no uso da interação por voz em demandas como aquela informada acima.

Neste contexto, alguns dispositivos oferecem a possibilidade para que seus usuários interajam através da voz, enquanto outros usam a voz como único meio de interação. Nos *smartphones* é possível observar sistemas que utilizam tanto o canal auditivo quando o visual em suas interações. Como exemplo disso, temos o *Google Assistente* (Figura 01) e a *Siri* (Figura 02), que indicam locais, vídeos, fotos, *links* e outros objetos de forma visual para seus usuários, mesmo que o *input* seja através da voz, como podemos ver nas figuras a seguir:



Figura 01- Google Assistente. Fonte: autor

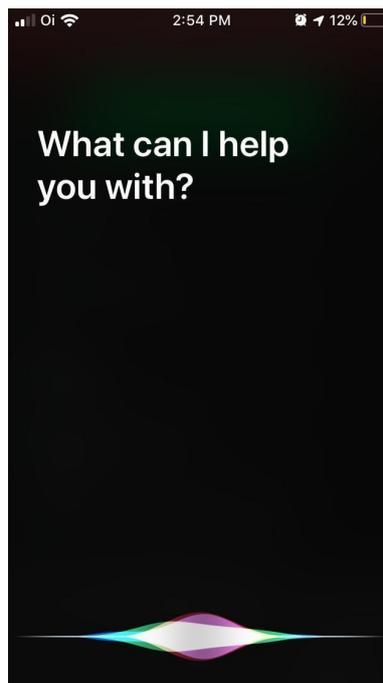


Figura 02 – Siri. Fonte: autor

Como mencionado anteriormente, existem dispositivos para os quais a voz é a única forma de interação com seus sistemas, tais como *Amazon Echo* (Figura 03) e o *Google Home* (Figura 04), que são os alto-falantes inteligentes (também conhecidos como assistente de casa). A diferença entre os dois modelos de interface não é de fácil compreensão, já que os dois modelos possuem muitas particularidades em comum. Assistentes Virtuais Inteligentes geralmente são acompanhados de recursos visuais, dado que normalmente são utilizados através de *smartphones*. Contudo, esse tipo de recurso não é restrito a este modelo de *hardware*, dado que alguns alto-falantes inteligentes atualmente possuem telas, o que possibilita que usuários utilizem tanto os canais auditivos quanto os visuais.

Segundo o *European Data Protection Supervisor* (2019),¹ alto-falantes inteligentes são porta-vozes que possuem na sua estrutura física (*hardware*) microfones que possibilitam interações por parte dos usuários, com outros dispositivos ou serviços. Já o Assistente Virtual Inteligente, é definido como o "cérebro" que transforma um alto-falante comum num serviço de *software* capaz de receber a voz do usuário, identificar o que foi falado, interagir com serviços e fornecer respostas.

¹ Autoridade Europeia para Proteção de Dados.

Figura 03- Amazon Echo ²Figura 04 - Google Home ³

Entretanto, o aumento da venda e uso de sistemas baseados em interação por voz não está associado unicamente ao sucesso desse modelo de interação já que, segundo Ben Sauer – designer de produto atuante na área de interface de voz e design de conversação –, em entrevista ao UX Podcast (2019), muitos usuários tendem a se desapontar com os resultados finais. Isso porque há uma grande expectativa sobre as interações, pois para os usuários não é possível visualizar a complexidade que há por trás de modelos de interação baseados em conversação.

É possível observarmos situações que promovem tais descontentamentos no dia a dia dos usuários que utilizam esse modelo de sistema. Podemos presenciar situações como essas sendo representadas até nas produções cinematográficas mais recentes, como no filme *Nós* (2019), do diretor Jordan Peele (Figuras 05 e 06).

² Disponível em <https://www.amazon.com.br/>. Acesso em 29 jul.20

³ Disponível em https://store.google.com/product/google_nest_mini Acesso em 29 jul.20



Figura 05 – Cena do filme *Nós* (2019), direção de Jordan Peele.



Figura 06 – Cena do filme *Nós* (2019), direção de Jordan Peele.

Não podemos afirmar que cenas como essas são comuns. Entretanto, situações análogas a essas, respaldadas por erros de compreensão por parte de sistemas baseados em interação por voz, atualmente são comuns a muitos usuários.

1.2

Objetivos, roteiro metodológico e justificativa da investigação

Sabemos que sistemas baseados em interação por voz têm ganhado cada vez mais espaço no mercado e na vida das pessoas e que ainda há muito a ser pesquisado

em torno desses sistemas. Portanto, este estudo visa contribuir com discussões em torno de sistemas baseados em interação por voz - também conhecidos pelo mercado como interfaces conversacionais ou interfaces de voz, fomentando-as.

O objetivo que norteia essa dissertação é a identificação de frustrações e limitações provenientes do uso de interfaces por voz no que se entende atualmente como conversa entre o humano e a máquina, relacionando-as de forma categórica à estrutura conversacional decorrente da Teoria da Conversação.

A justificativa para a realização desta pesquisa apoia-se na observação de que há pouco material sobre o assunto em nossa língua materna - o português. Esse material torna-se ainda mais escasso ao aplicarmos um recorte racial em torno de quem produz pesquisas científicas sobre o tema já que ao longo deste trabalho foram encontrados e utilizados um grande número de autores brancos.

Ainda no aspecto social esta pesquisa também se torna relevante quando propomos a compreensão do problema aqui descrito. Quando observamos as dificuldades encontradas pelos usuários podemos então propor soluções ou caminhos capazes de no futuro solucionar questões as quais atualmente não conseguimos, isso é claro, buscando facilitar a forma com a qual usuários interagem através da voz.

Se construirmos um cenário hipotético no qual uma pessoa cega, por exemplo, vive a 20 quilômetros da agência bancária mais próxima e precisa depositar alguma quantia para um outro alguém que mora em outra cidade podemos nos perguntar: qual seria a forma mais fácil para que ela conseguisse realizar a transação?

Quando olhamos para este cenário hipotético podemos nos permitir elaborar algumas respostas. (1) a pessoa privada de visão pode andar por 20 quilômetros até a agência; (2) a pessoa privada de visão pode acessar um aplicativo bancário e realizar a transação após passar pela curva de aprendizado necessária; (3) a pessoa pode pedir através da fala que um determinado valor saia da sua conta e vá para a conta de destino.

Quando olhamos essas três respostas é possível assumir que a forma mais fácil seria que a pessoa cega pudesse transferir um valor apenas pedindo através da voz, algo que ela já está habituada e a qual não possui limitações. Se olharmos para esta situação hipotética, podemos aferir que o estudo sobre a interação através da voz pode permitir desdobramentos capazes de facilitar a vida dos seres humanos

cada vez mais, democratizando acessos a informações e ações a todo ser humano que consiga se comunicar através da fala.

Esta pesquisa também se torna importante ao observamos usuários interagindo com esses sistemas, visto que é comum repararmos em interações interrompidas ou não concluídas com êxito. Tamanha é a naturalidade dessas interações - interrompidas ou não concluídas com sucesso - que mesmo sendo uma tecnologia difundida recentemente, já é possível encontrar representações dessas situações nas grandes produções cinematográficas, como retrata o filme *Nós* (2019) do diretor Jordan Peele.

Assim, com base nos problemas identificados dessa interação do humano com esse tipo de interface, questionamos: sistemas baseados em interfaces de voz conseguem manter uma conversa com seres humanos para além do primeiro turno⁴?

A hipótese que orienta esta pesquisa considera que sistemas baseados em interação de voz, como estão projetados atualmente, não são capazes de manter uma conversação por mais de um turno conforme os preceitos da Teoria da Conversação.

Buscando responder de forma adequada a questão levantada, realizamos um levantamento do estado da arte dos conhecimentos produzidos em torno da comunicação humana e da área da ergonomia humano-máquina para que pudessem contribuir com informações em torno do universo que circunda as interações por voz.

A partir do levantamento realizado, mostrou-se necessário um aprofundamento na investigação, levando-nos a aplicar a técnica de entrevista como forma de coletar dados diretamente com os usuários do sistema pesquisado, resultando posteriormente na análise de conteúdo das respostas obtidas.

Observando as inferências provenientes da análise de conteúdo, elaboramos uma nova proposta de modelo interacional, que corresponde à forma com que se é dada atualmente à interação entre o humano e sistemas baseados em interface de voz.

Esta dissertação visa a trazer considerações gerais em torno das experiências no uso dos sistemas conversacionais atuais que possam apoiar futuros pesquisadores, alunos e interessados no tema.

⁴ Um turno é caracterizado pelo comprimento dos cinco pontos conceituais empregados na conversa

1.3

Estrutura da dissertação

Para viabilizar os objetivos propostos nesta investigação, a dissertação está estruturada em oito capítulos (incluindo esta introdução) permeados por fundamentação teórica, entrevistas, análises, resultados e conclusões.

No capítulo 2, discorremos sobre o olhar cronológico aplicado às formas empregadas à comunicação humana, considerando: as comunidades primitivas, as sociedades tradicionais, as sociedades complexas e a disseminação da *internet* como provedora de mudanças que cruzam a comunicação.

No capítulo 3, apresentamos referenciais teóricos capazes de definir o termo interface, objeto inerente às interações, iniciando o recorte dos fundamentos ergonômicos humano-máquina.

Já no capítulo 4, efetuamos o recorte e a abordagem dos fundamentos da ergonomia humano-máquina abrangendo: definição de interação, o papel das atividades mentais nas interações, o que contempla uma interação sonora, o que é uma interação por voz e a conversa como modelo de interação.

Por sua vez, no capítulo 5, apresentamos o delineamento dos procedimentos metodológicos aplicados, explicitando o planejamento e a condução de entrevistas como técnica de investigação.

No capítulo 6, apresentamos o planejamento e condução da técnica de análise de conteúdo como recurso para extração de significado das respostas obtidas com o uso da técnica entrevista.

Propomos, no capítulo 7, uma nova perspectiva sobre o modelo de interação aplicado aos sistemas baseados em interação por voz, segundo as inferências coletadas através do resultado da análise de conteúdo.

Por fim, no capítulo 8, tecemos as considerações em torno das experiências no uso de sistemas baseados em interação por voz, as conclusões finais extraídas da pesquisa e os desdobramentos futuros.

2

Sociedade e Comunicação

Neste capítulo, iremos analisar a história da comunicação humana abordando e explorando o que Martino (2005) define como comunidade primitiva, sociedade tradicional e sociedade complexa. Abordaremos também a chegada e a democratização da internet como forma de comunicação e como sua expansão afeta a forma com a qual estávamos acostumados a nos comunicar até então.

2.1

A comunidade primitiva

Para compreendermos a dinâmica por trás da comunicação entre o humano e seu assistente virtual inteligente, precisamos entender a trajetória dessa relação.

Precisamos olhar para a história e analisá-la buscando compreender as relações do ser humano com a comunicação e também com a tecnologia, explorar os conceitos, as definições e os simbolismos que serão capazes de elucidar a forma com a qual lidamos com essa relação que se consolida cada vez mais através dos avanços tecnológicos.

Neste capítulo, portanto, buscamos olhar para o passado, analisando o vínculo do indivíduo com as formas de se comunicar desde os tempos remotos até a atualidade, atravessando, não obstante, a utilização da técnica na comunicação.

Durante a evolução humana, desde as comunidades primitivas até a construção da vida complexa, o ser humano, por necessidade, criou formas de sobrevivência. Reproduziu o fogo, inventou a roda, aprendeu a cultivar, produziu conhecimentos sobre o planeta em que habita, investigou o espaço, entendeu como o seu próprio corpo funciona: tudo através da técnica. E com o modo de se comunicar não poderia ter sido diferente.

Segundo Martino (2005, p. 1), a linguagem não foi apenas uma questão para o ser humano, mas também uma forma de compreender o que é o humano, uma vez

que, segundo o autor, existem pelo menos, duas características que separam o ser humano dos outros animais: a linguagem e a técnica. É através delas que antropólogos conseguiram definir a espécie humana, pontuando que "é somente lá, onde há restos de atividade técnica e de atividade simbólica, que podemos distinguir o homem de outros primatas" (Martino, 2005, p. 1).

Tendo em vista a fala como uma das formas de linguagem, voltamos nossa atenção para a comunicação – a qual ocorre através da linguagem, capacidade que pertence apenas ao ser humano, como definido por Hohlfedt, Martino & França (2008) –. Os autores destacam que o ser humano é um ser eminentemente social, incapaz de viver isolado e solitário, determinando, assim, que a comunicação é também um fenômeno de cunho social.

A partir da compreensão de que a comunicação é um fator necessário para o entendimento do que é ser social, resgatamos a história de configuração do modo de vida em sociedade desde seus primórdios. De forma cronológica, Martino (2005) estrutura o perfil comunicacional de três tipos de sociedades que, sequencialmente, demonstram os avanços comunicacionais através dos tempos, permeando inclusive a "intervenção da técnica no domínio da comunicação" (Martino, 2005, p. 1).

Na primeira instância dessa estrutura cronológica, encontra-se a formação do que Martino (2005) chama de comunidade primitiva. Nessa formação social destaca-se o estilo de vida dos indivíduos, característico de atividades como a pesca, a caça e o extrativismo, todas baseadas nos recursos naturais disponíveis em seus territórios. É importante ressaltar que a comunidade primitiva é composta por pequenos grupos contendo algumas centenas de pessoas que, devido às limitações espaciais, acaba resultando numa baixa densidade demográfica. Nesses grupos a arquitetura comunicacional é essencialmente oral (Martino, 2005, p. 2).

Este contexto demográfico delimitado até mesmo pela restrição da marcha humana torna a repartição do conhecimento técnico homogêneo, já que as vivências e experiências de indivíduos de uma determinada comunidade constituíam um intensivo saber, por exemplo, sobre técnicas específicas. A comunicação interpessoal se mostra suficiente para o compartilhamento da experiência coletiva:

O compartilhamento extensivo e intensivo de uma mesma experiência de vida, que somente a comunidade primitiva pode proporcionar, faz com que a comunicação interpessoal seja plenamente suficiente para este tipo de organização social. (Martino, 2005, p. 2)

De acordo com Lima (2007), os indivíduos pertencentes a uma comunidade estritamente oral, como as comunidades primitivas, detinham apenas como recurso o uso da memória, que auxiliava "ao longo do tempo, reter e transmitir as representações que lhes eram convenientes de perdurar" (Lima, 2007, p. 276).

Neste cenário, ainda segundo esse autor, recursos como a dramatização, personalização e artifícios narrativos eram aplicados para que houvesse uma maior possibilidade de prolongar, dentro das comunidades, os conhecimentos adquiridos através dos tempos, dado que a única possibilidade de haver trocas de mensagens linguísticas entre indivíduos se dava pela necessidade de que ambos, emissor e receptor, estivessem no mesmo contexto (tempo e lugar), e universo semântico.

Não havia, nessa conjuntura, uma forma de registrar a realidade, nenhuma forma sistêmica, pois tudo era armazenado através da memória e, com o "transcorrer do tempo, exigia um contínuo recomeço, uma renovação suscetível a alterações visíveis de geração para geração" (Lima, 2007, p. 276). No entanto, havia formas de facilitar o registro e a transmissão da memória.

Rituais mágico-religiosos estão intrinsecamente ligados à necessidade de perdurar conhecimento como, por exemplo, nas lendas, nos mitos, nos cantos e nos poemas, sendo esses realizados estritamente por tradição oral.

2.2

A sociedade tradicional

Três fatores nos levaram ao surgimento da denominada sociedade tradicional, segundo Martino (2005). O primeiro foi a existência da agricultura, que culminou nas chamadas sociedades sedentárias, já que por meio dessas houve um aumento significativo da produção alimentícia. O segundo fator foi o surgimento de grandes concentrações urbanas; e, por fim, o advento da escrita, importante desenvolvimento técnico do ser humano na forma de se comunicar. (Martino, 2005, p. 3)

Martino (2005) associa ainda o surgimento da cidade com as condições sociais provenientes da agricultura, o que por sua vez culmina no desenvolvimento do Estado que estabelece a escrita como um "instrumento de administração das

riquezas trazidas pela agricultura e organização político-social da cidade" (Martino, 2005, p. 3).

Da mesma forma com que a fala teve sua importância como principal instrumento de uma era, cunhada por Lima (2007) como "tempo da oralidade primária" (Lima, 2007, p. 276-277), a escrita, próxima da que conhecemos hoje - com ideogramas e equivalências fonéticas -, surgiu na Mesopotâmia, por volta de 3.000 anos a.C. Segundo a autora, no antigo Egito já se fazia uso de papiros e tintas rudimentares para elaboração de documentos. Tempos depois, no século III a.C., outro formato surgiu como opção de suporte, o pergaminho.

Lima (2007) descreve que no início do século III a.C. surgiu a Biblioteca do Museu de Alexandria, que tinha o objetivo de reunir todo o conhecimento humano em um só lugar através de livros que eram formados por papiros e pergaminhos, os quais se tornaram um objeto de poder, conforme é apresentado no seguinte trecho: "A partir desse momento, o livro passou a objeto de autoridade e prestígio, tornando-se sinônimo do saber" (Lima, 2007. p. 277).

Uma grande invenção vinculada à escrita é o alfabeto que, segundo Diringir (1968, *apud* Lima, 2007, p. 277), é uma das maiores invenções do homem, já que foi a forma de se escrever altamente desenvolvida e mais conveniente, sendo o sistema mais adaptável existente:

A invenção do alfabeto não somente permitiu à humanidade comunicar ideias por símbolos visuais, mas também a criação de registros permanentes destes signos e, assim, a criação de uma memória externa à mente humana. (Diringir, 1968, p. 14, *apud* Lima, 2007. p. 277).

A escrita representa a complexificação da vida social. A partir dela denotam-se diferentes papéis para os indivíduos de uma sociedade, culminando, assim, na diferenciação desses indivíduos por *status* sociais. Como técnica, a escrita era tida como uma forma de organização cultural, destinada apenas ao clero e à nobreza, classes definidas como superiores. A transmissão de aprendizados e experiências não se dava unicamente através da fala, mas também com novos recursos como o livro, que "exterioriza a memória, tirando-a do registro psicológico para torná-la objetiva, material" (Martino, 2005, p. 4).

A escrita se refletia na divisão de classes, como bem reforça Martino (2005). Apesar de seu advento, os mais pobres, como os servos e as pessoas escravizadas

mantinham-se informados apenas pela tradição oral, já que os grupos considerados superiores contavam com o que o autor chama de “cultura do letramento” (Martino, 2006, p. 5). Assim, qualquer classe que não pertencesse a esses grupos foi privada da utilização da técnica utilizada para perdurar informações, não excluindo os costumes culturais, ou seja, "a escrita, para a classe dominante, e a tradição oral para a classe dominada" (Martino, 2006, p. 5).

Sob outra perspectiva, a escrita veio para desvincular a necessidade em se ter dois indivíduos, emissor e receptor, presentes no mesmo tempo e lugar; e sem requerer a experiência decorrente do mesmo universo semântico, desassociando o discurso da situação na qual foi produzido:

Se o seu registro escrito “fala por si mesmo” sofre, por outro lado, interferência de quem o “consulta”. Conceitualmente, o tempo e o espaço tornam-se diferenciados daquilo que são no universo da oralidade, em que as adaptações de contexto e tempo são eliminadas. Há uma autonomia do texto em relação à tradição oral. O tempo não é mais o da circularidade, mas sim da linearidade, linearidade que se traduz no transcorrer da História. Qualquer discurso torna-se possível de ser apreendido, analisado e interpretado fora de seu contexto de produção. A escrita cria a figura do leitor, para o qual a realidade passa por um filtro muito mais refinado do que ocorre com o ouvinte no tempo da oralidade. (Lima, 2007, p. 278)

Neste período a escrita ganha força e assume outro lugar por meio de Johannes Gutenberg que, mesmo não sendo o único a pesquisar sobre impressão de textos, foi o que criou o modelo europeu de impressão trazendo para o Ocidente o que já era feito a pelo menos um milênio pelos chineses. No século XV, Gutenberg criou um tipo móvel de metal que, junto a outros tipos móveis, era capaz de registrar palavras e frases em uma mesma linha de forma ordenada (Rodrigues, 2012).

Além das facilidades provenientes desta invenção, não podemos deixar de evidenciar a importância que isso teve na desconstrução e diminuição do que a própria escrita criou: a segregação das classes.

Consoante aos pensamentos de Rodrigues (2012), observamos que as cópias manuscritas tinham um custo econômico muito grande e eram, escrupulosamente, controladas por grandes grupos de copistas, pela Igreja e pelo Estado. Ou seja, se tornavam acessíveis apenas para indivíduos privilegiados: ricos e influentes.

Nesse contexto, é importante entender a conotação social que a escrita estabelecia até o momento, pois, através da invenção de Gutenberg, barreiras sociais em torno da comunicação começaram a ser quebradas. Este fato nos permite fazer

uma analogia com o que iremos discutir mais à frente, neste capítulo, com a chegada da internet no século XX. Ainda em consonância com o autor, pode-se dizer que:

O livro impresso representou não apenas uma opção de qualidade e mais acessível comercialmente, como também aumentou incrivelmente a liberdade da sociedade com relação à leitura; foram afrouxadas as proibições devido à impossibilidade de controlar uma demanda crescente por livros de todos os tipos. (Rodrigues, 2012, p.190)

2.3

A sociedade complexa

Por mais que a linguagem verbal fizesse parte da sociedade tradicional, a forma de comunicação mais prestigiada neste contexto era a escrita, o que excluía qualquer indivíduo ou instituição que não estivessem relacionados à religião ou ao estado. Na composição da sociedade complexa é possível constatar o que Martino (2005) chama de "uma verdadeira explosão comunicacional" (Martino, 2005, p. 5).

Os meios de comunicação, segundo Martino (2005, p. 5), diversificaram-se com a criação de ferramentas como o rádio, a TV e o telefone. Mas não apenas ficaram diversificados, como também passaram a encontrar-se em abundância e ao alcance de uma parcela populacional maior e a serviço do indivíduo, diferentemente da escrita, no contexto da sociedade tradicional.

É na sociedade complexa que aparecem os primeiros indícios da comunicação oral humano-máquina, tema esse que é oportuno para esta pesquisa de mestrado. Na década de 1950, os pesquisadores K. H. Davis, R. Biddulph e S. Balashek criaram um dos primeiros sistemas para reconhecimento de dígitos, chamado de Audrey (Warren, 2014). Esse sistema era capaz de reconhecer comandos de um único falante na identificação de dígitos de 0 a 9. Audrey - como precedente dos sistemas de comunicação oral que temos atualmente - tornou-se, desta maneira, o marco zero do advento desta tecnologia, o qual abordaremos no segundo capítulo dessa dissertação, com uma discussão mais aprofundada acerca da evolução do sistema (Warren, 2014).

O novo contexto que emergiu da pluralização dos meios de comunicação tem sua verdadeira significação devido a duas características marcantes na sociedade complexa: a primeira é o nível de complexidade a respeito da organização social desse contexto, que consegue ultrapassar as limitações espaciais em consequência

da explosão demográfica, do aparecimento de concentrações urbanas extremamente populosas, do desenvolvimento de uma economia de mercado e da globalização e sua totalidade. Foi instituída, desse modo, uma demanda comunicacional nos âmbitos sociais, econômicos e culturais (Martino, 2005, p. 5).

A segunda característica é o movimento que dá ao indivíduo uma posição de emergência e que é decorrente, em grande escala, do âmbito político que define o indivíduo como eleitor e cidadão; do âmbito econômico que classifica o indivíduo em pelo menos três classes: trabalhador, empresário e investidor; e também no âmbito social que relativamente liberta o indivíduo do coletivo (Martino, 2005, p. 5).

Enquanto indivíduo pertencente à sociedade complexa, sua inclusão cultural e social depende agora, além dos fatores externos a ele, da sua própria capacidade em se firmar em espaços profissionais, em seus grupos sociais, da construção dos seus valores na compreensão e vivência de sua cultura. É nesse novo contexto que se estabelece a importância indiscutível que a comunicação exerce na organização social:

É isto que está por trás da grande explosão da comunicação que assistimos no século XX, quando a lenta fermentação de certas condições sócio históricas (algumas de longo termo) finalmente se junta com as novas condições tecnológicas dos modernos meios de comunicação. (Martino, 2005, p. 7)

Essa “junção” mencionada acima por Martino (2005) resulta em uma nova arquitetura comunicacional, interferindo inclusive nas concepções que o indivíduo tem sobre cultura e revolucionando a forma com a qual a experienciamos.

Desse modo, existem para a sociedade complexa as condições para que o indivíduo compartilhe suas experiências sociais através da virtualização da técnica. "Agora a comunicação se apresenta transfigurada em fenômeno técnico e passa a ser tomada como centro de referência para a vida social" (Martino, 2005, p. 8).

2.4

A explosão da *internet* como meio comunicacional

Como vimos, é no contexto da sociedade complexa - seguindo a definição de Martino (2005) - que ocorre o surgimento de novas formas de mídias, em especial

a televisão. Isso culmina no que Castells (1999, p.417) chama de o “fim da galáxia de Gutenberg” isto é, o fim de um sistema comunicacional essencialmente dominado pela natureza tipográfica e pela ordem fonética do alfabeto. Não muito tempo depois do fim da galáxia de Gutenberg, uma nova forma de se comunicar ganhou ainda mais força em meio aos já tradicionais rádios e TVs: a internet.

Inicialmente planejada para uso militar, a internet começa a surgir nas últimas décadas do século XX. De acordo com Lemos (2010, p. 102), pesquisadores como Licklider encorajaram a ideia da formação da rede até então conhecida pelo seu nome de batismo: *ARPANET*⁵, que traduzido significa Rede da Agência para Projetos de Pesquisa Avançada.

O objetivo da rede até aquele momento era manter o tráfego de informações a salvo de qualquer possível interrupção, caso fosse instaurada uma guerra nuclear decorrente da Guerra Fria:

Com base na tecnologia de comunicação por comutação de pacotes, o sistema tornou a rede independente de centros de comandos e controle, de modo que as unidades de mensagens encontrassem suas rotas, sendo remontadas com sentido coerente em qualquer ponto do sistema. (Ferreira de Sá, 2005 *apud* OLIVEIRA, 2007, p. 31).

Em sua essência, podemos afirmar que nesse sistema a informação era separada em fragmentos e cada fragmento era transportado de maneira independente, de forma que cada um deles fosse indecifrável sozinho. Todos os fragmentos eram transportados com o objetivo comum de chegar ao mesmo ponto, sendo esta a única maneira de remontar à informação inicial para, então, tornar possível a obtenção de sentido.

Segundo Oliveira (2007, p. 31), só no ano de 1992 a tecnologia digital conseguiu alcançar um patamar de desenvolvimento superior. A partir desse ano, passou a ser possível a transmissão de imagens e sons por essa rede. A autora descreve que estudantes da Universidade de Illinois desenvolveram um *software* capaz de permitir a navegação entre redes, o que possibilitou a criação da internet como a conhecemos atualmente, a *World Wide Web*.

É importante elencar aqui a internet com o desenvolvimento da microinformática, que surge na metade dos anos de 1970 (Lemos, 2010, p. 103) -

⁵ *Advanced Research Projects Agency Network*

essa que mais tarde resultou na criação de plataformas capazes de dar acesso à rede. Ainda segundo Lemos (2010, p. 99), o surgimento da microinformática está relacionado diretamente com outros domínios científicos, como a cibernética (1948), a inteligência artificial (1956) e a teoria da auto-organização e de sistemas da década de 1960.

É apenas no ano de 1975 que surge um dos primeiros microcomputadores, o *Altair*. Em seguida vem o *Apple II*, em 1977 e, posteriormente, em 1981 e 1984, são criados, respectivamente, o primeiro PC (*personal computer*), baseado em um modelo da IBM, e o *Apple Macintosh* (Lemos, 2010, p. 103).

A criação do padrão agora definido pelo *Apple Macintosh*, com vistas à democratização da microinformática, possibilitou o avanço da área trazendo certas discussões:

A democratização dos computadores vai trazer à tona a discussão sobre os desafios da informatização das sociedades contemporâneas, já que esses não só devem servir como máquina de calcular e de ordenar, mas também como ferramentas de criação, prazer e comunicação; como ferramenta de convívio. (Lemos, 2010, p. 104)

O desenvolvimento dos microcomputadores possibilitou que a internet se tornasse a espinha dorsal da comunicação global dos anos 1990 (Castells, 1999). Nessa década a internet era acessada por aproximadamente 25 milhões de usuários e com a possibilidade de enviar qualquer tipo de mensagem, incluindo imagens e áudios. Com isso, foi possível a criação de uma rede que conseguia transmitir toda e qualquer variedade de símbolos.

Nesse período, o mercado mostrou, pela primeira vez, um forte interesse em se conectar por meio da internet, graças à decisão tomada pela *National Science Foundation*⁶ em privatizar parte de suas operadoras. A partir disso, surgiu uma comercialização que se expandia em velocidade acima do esperado, conectando diversas redes comerciais de serviços.

O governo e o mercado americanos, possuíam, então, um novo interesse em comum: a *World Wide Web*. Com a internet presente na vida de milhões de seres humanos, a estrutura comunicacional obteve uma grande expansão e novas formas de comunicação surgiram a partir dela. Um exemplo disso é a comunicação oral

⁶ Fundação nacional de ciência americana

humano-máquina que, por mais que existam estudos datados desde 1950, nos dias de hoje ainda vem se ampliando e gerando discussões.

É com referência a essa nova estrutura de comunicação que esta pesquisa pretende se debruçar, abordando a forma com a qual seres humanos conseguem se comunicar com máquinas e realizar tarefas que facilitem o seu dia a dia por meio de ações eficazes, tendo como recorte a comunicação oral entre indivíduo e objeto através da interface do usuário por voz.

Como ponto de partida, iremos observar o uso do som na realização de interações entre o humano e a máquina, conforme demonstrado no capítulo a seguir.

3

Design de interação em sistemas conversacionais

Para tratarmos de interação, seja ela auditiva ou não, precisamos primeiro entender o conceito e a definição que está por trás do termo interface, já que é através dele que as interações se tornam possíveis.

3.1

O que é uma interface?

O ser humano está acostumado a lidar com interfaces o tempo todo. Vejamos: em uma conversa com um grupo de amigos? Sim, há interface ali. Pedindo um hambúrguer em um totem? Também há interface ali. Muitos acreditam que a interface tenha nascido junto com a grande explosão de vendas de microcomputadores, mas na verdade interfaces existem desde os tempos mais primórdios, quando a única forma de comunicação era ainda a linguagem não falada.

Estamos acostumados a lidar com interfaces quando usamos um computador, um *smartphone*, um *smartwatch* ou quando interagimos através de nossas redes sociais. Mas, se a interface não é apenas o que vemos quando interagimos através do *Instagram*, o que é uma interface?

Segundo Moran (1981 *apud* Barbosa e Silva, 2010), uma interface que permite a interação com um determinado sistema é toda a parte que possibilita o contato físico, seja ele motor ou perceptivo, ou contato conceitual durante toda a interação, sendo o único meio ao qual é permitido ao usuário interagir com o sistema. Por isso, por muitas vezes é possível perceber que o usuário passa a confundir a interface com o próprio sistema, já que a interface é o que está em contato direto com o usuário como citado por Hix & Hartson (1993 *apud* Barbosa e Silva, 2010).

Para Hugo (1995, p. 8), uma interface é a parte do sistema que se encarrega de produzir um limite comum para os usuários. Entendendo que, para que haja comunicação, o sistema deva se assegurar do que o autor nomeia de conexão física entre as entidades. Essa conexão física deve traduzir as formalidades de cada linguagem envolvida no ato da interação, sendo essa a única forma de se estabelecer uma comunicação entre ambas as partes. Thro (1991 *apud* Hugo, 1995, p. 8) afirma que as interfaces servem como um local de encontro ou interação entre o humano e o sistema.

Barbosa e Silva (2010), utilizando-se da definição de Moran (1981), explicitam o que compreendem sobre os dois tipos de contatos possíveis com uma interface, definindo para eles como ocorre o contato físico:

O contato físico na interface ocorre através do hardware e do software utilizados durante a interação. Dispositivos de entrada, como teclado, mouse, joystick, microfone, caneta (que escreve sobre a tela) e câmera (webcam), permitem ao usuário agir sobre a interface do sistema e participar ativamente da interação. Já os dispositivos de saída, como monitor, impressora e alto-falante, permitem ao usuário perceber as reações do sistema e participar passivamente da interação (Barbosa e Silva, 2010. p. 25).

Da mesma forma, os autores também definem o contato conceitual:

O contato conceitual com a interface envolve a interpretação do usuário daquilo que ele percebe através do contato físico com os dispositivos de entrada e de saída durante o uso do sistema. Essa interpretação permite ao usuário compreender as respostas do sistema e planejar os próximos caminhos de interação (Barbosa e Silva, 2010. p. 26).

Para esses estudiosos, a interface é o que determina quais são as possibilidades direcionadas aos usuários dentro de um sistema de acordo com o que ele pode falar ou fazer. Para eles, a partir do momento em que se é definido o tipo de interação de um sistema e como ela deve ocorrer, estamos então, restringindo certas características da interface. Barbosa e Silva (2010, p. 26) pontuam também a necessidade de levar em consideração o contexto dos usuários para os quais uma interface está sendo projetada: não adianta desenhar uma interface com um passo-a-passo escrito para um usuário analfabeto.

Podemos presumir, por todo o exposto, que uma interface é um lugar comum de comunicação entre humano e máquina, o qual se encarrega de traduzir a

linguagem empregada por cada uma das partes, a fim de possibilitar a troca de informação entre elas tendo, como resultado, o que chamamos de interação.

3.2

Do termo *design* de interação à definição de interação

O termo *design* de interação foi proposto pelo diretor da empresa de *design* IDEO, Bill Moggridge, em meio aos anos de 1990. Moggridge percebeu que ele e alguns colegas utilizavam-se de algumas ferramentas de distintas disciplinas, para a concepção de seus produtos, como por exemplo, as do campo do *design* de comunicação, e de conceitos da ciência da computação, já que muitos desses termos se relacionava à criação de seus produtos com computadores e *softwares*. A nomenclatura *design* de produtos parecia insuficiente, portanto, para descrever o que se era realizado por eles, por mais que naquela esfera ainda se desenvolvessem produtos (Saffer, 2010, p. 2).

Para ele, todas essas disciplinas que faziam parte de seus processos de forma perceptível, englobavam o que ele passou a chamar de *design* de interação, já que, além de seus processos percorrerem diferentes disciplinas, também havia, por trás do palpável, o conceito de conectar as pessoas através dos produtos utilizados por ela (Saffer, 2010, p. 2). Esse autor define muito bem o que é e de onde se oriunda o termo *design* de interação. Mas, além desse termo específico, precisamos entender o que é ou o que está sob o termo interação.

3.2.1

O que é interação?

Para Barbosa e Silva (2010, p. 20), o termo interação vem evoluindo bastante ao longo do tempo. Segundo os autores, esse termo trazia na sua originalidade a tratativa de sequência a estímulos e respostas, a exemplo de interação entre corpos físicos. A compreensão do termo evolui por meio do surgimento de pesquisas fundamentadas na área da cognição, as quais passam a enfatizar a interação como a comunicação do humano com a máquina. A interação também pode ser definida como o processo no qual o homem, ao interagir com a máquina, formula uma intenção, planeja suas ações, atua sobre a interface, percebe e interpreta a resposta do sistema, avaliando, por fim, se seu objetivo foi alcançado (Barbosa e Silva, 2010, p. 20).

Esses autores - baseando-se nas ideias de Hix e Hartson (1993) - trazem outras definições para o termo interação, a saber: "a interação usuário-sistema pode ser considerada como tudo o que acontece quando uma pessoa e um sistema computacional se unem para realizar tarefas, visando a um objetivo." (Barbosa e Silva, 2010, p. 20). E a segunda, ainda mais recente, proposta por Souza (2005 *apud* Barbosa e Silva, 2010, p. 20), descreve o termo interação (usuário-sistema/homem-máquina) como: "processo de comunicação entre pessoas, mediada por sistemas computacionais".

Por fim, os autores ainda nos trazem suas considerações sobre a definição do termo interação e a classificam como sendo "um processo de manipulação, comunicação, conversa, troca, influência e assim por diante" (Barbosa e Silva, 2010, p. 20).

Ainda na intenção de percorrer e investigar as definições de interação, podemos analisar um artigo elaborado por três pensadores que caminham entre diversas disciplinas, dentre as quais estão o *design* e a cibernética. Dubberly *et al.* (2009) trazem em seu artigo diferentes perspectivas sobre o que é o termo interação de acordo com algumas disciplinas. Os autores dividem o texto em três pontos de vista, que são sobre a perspectiva do *design*, a perspectiva da interação humano-computador (IHC), e a perspectiva que eles denominam de sistemas (2009, p. 1-3).

Na perspectiva do *design*, os autores apontam o argumento de Meredith Davis⁷ de que interação não é uma disciplina voltada apenas para produtos relacionados aos computadores, ressaltando que, por exemplo, livros impressos também são pensados pelos *designers* como uma forma de interação, já que esses profissionais consideram as formas com as quais os leitores irão interagir com as plataformas físicas. Davis cita, como exemplo, os trabalhos de Massimo Vignelli para a *National Audubon Society*⁸, especialmente o *Field Guide to North American Birds*⁹, o qual considera cuidadosamente pensado para a interação.

Dubberly *et al.* (2009) reafirmando a perspectiva de Davis (2008) destacam o pensamento de Richard Buchanan, o qual contrasta a estrutura do *design* - que antes mantinha seu foco na forma e, posteriormente, em torno dos significados e dos contextos - com uma estrutura ainda mais atual, na qual ele define o foco, agora, na interação:

A interação é uma maneira de enquadrar a relação entre pessoas e objetos projetados para elas - e, portanto, uma maneira de estruturar a atividade do design. Todos os objetos criados pelo homem oferecem a possibilidade de interação, e todas as atividades de design podem ser vistas como design para interação. O mesmo acontece não apenas com objetos, mas também com espaços, mensagens e sistemas. A interação é um aspecto essencial da função e a função é um aspecto essencial do design. (Buchanan, 1998 *apud* Dubberly *et al.* 2009. p.1, tradução nossa¹⁰)

Observando o que os autores acima chamam de perspectiva de IHC, temos uma bem próxima daquelas apresentadas anteriormente por Barbosa e Silva (2010). Dubberly *et al.* (2009) trazem a ideia da interação conforme apresentada pelos modelos canônicos baseados na estrutura arquetípica - conhecida como ciclo de *feedback* - estrutura essa que interfere no fluxo de informações entre um sistema e uma pessoa seguindo etapas as quais:

⁷ Meredith Davis é educadora, escritora e *designer* gráfica. Seu trabalho se concentra em defender uma educação abrangente, crítica e desafiadora em *design*.

⁸ A *National Audubon Society* é uma organização não governamental de conservação da natureza, fundada em 1905 na cidade norte-americana de New York.

⁹ Guia de Campo para Aves da América do Norte.

¹⁰ "Interaction is a way of framing the relationship between people and objects designed for them—and thus a way of framing the activity of design. All man-made objects offer the possibility for interaction, and all design activities can be viewed as design for interaction. The same is true not only of objects but also of spaces, messages, and systems. Interaction is a key aspect of function, and function is a key aspect of design."

A pessoa tem um objetivo; ela atua para alcançá-lo em um ambiente¹¹ (fornece informações para o sistema); ela mede o efeito de sua ação no ambiente (interpreta a saída do sistema - *feedback*) e compara resultado com objetivo. A comparação (produzindo diferença ou congruência) direciona sua próxima ação, iniciando o ciclo novamente. (Dubberly *et. al.*, 2009. p. 1, tradução nossa¹²)

Podemos observar melhor a definição do modelo de *feedback* na Figura 07, a seguir:

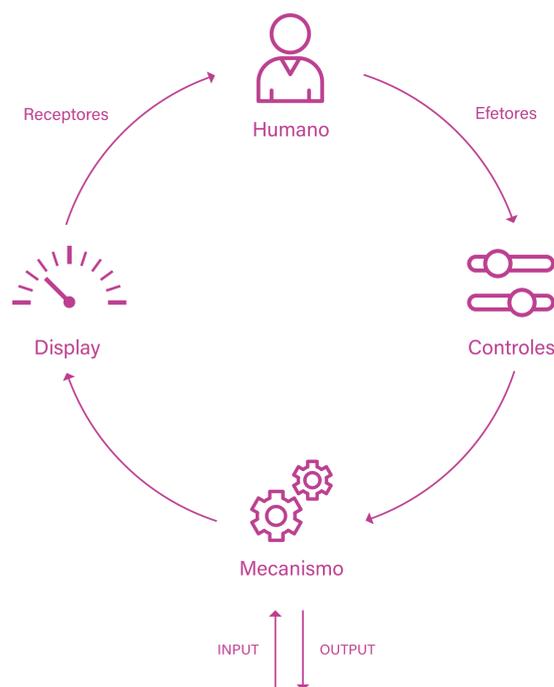


Figura 07 - Modelo de *Feedback* (Fonte: Dubberly *et al.*, 2009, p. 2).

Estas etapas formulam o que Dubberly *et al.* (2009, p. 2) chamam de sistema simples de autocorreção e categorizam-no como um sistema pertencente à primeira ordem da cibernética. (Como não é nosso objetivo explorar o que se refere à primeira e à segunda ordem da cibernética, seguiremos com outros exemplos.)

Norman (2002 *apud* Dubberly *et al.*, 2009, p. 1) propõe um modelo chamado por ele de "modelo do abismo", o qual descreve que há um "abismo" de execução e um "abismo" de avaliação entre o usuário e a plataforma física que comporta o

¹¹ Nesta pesquisa iremos tratar a tradução de ambiente como interface.

¹² "The person has a goal; she acts to achieve it in an environment (provides input to the system); she measures the effect of her action on the environment (interprets output from the system—*feedback*) and then compares result with goal. The comparison (yielding difference or congruence) directs her next action, beginning the cycle again."

sistema com o qual ele vai interagir. Dessa maneira, o usuário transforma sua intenção em ação. Acionando o sistema através de *input* em sua plataforma física, o sistema por sua vez responde ao usuário que interpreta e avalia a resposta, como podemos analisar na Figura 08.

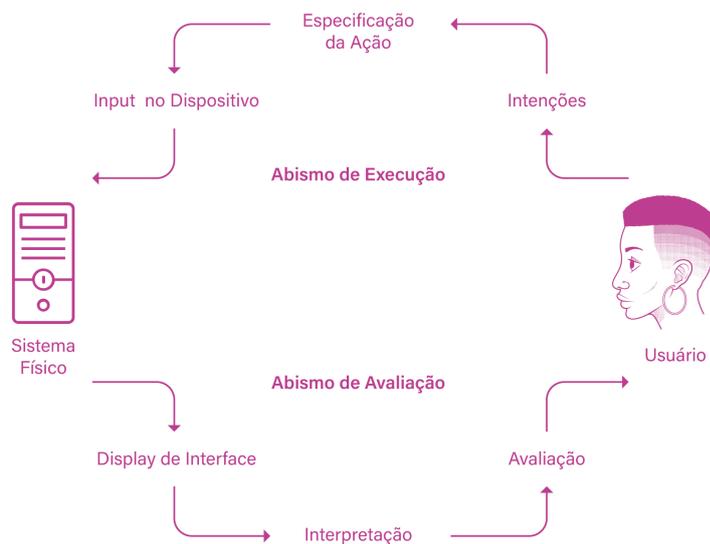


Figura 08 - Modelo de Abismo (Fonte: Dubberly, Pangaro e Haque, 2009, p. 2)

Norman (2008 *apud* Dubberly et al., 2009, p. 1) propõe uma variação deste modelo denominada de "sete estágios de ação", dando continuidade aos estudos que compuseram o modelo do abismo, conforme a Figura 09. Nessa variação, o autor ressalta que nem todo comportamento começa com um objetivo, descrito por ele como comportamento de cima para baixo, mas que também pode começar de baixo para cima, que se configura a partir de ações tomadas como respostas ao meio ambiente.

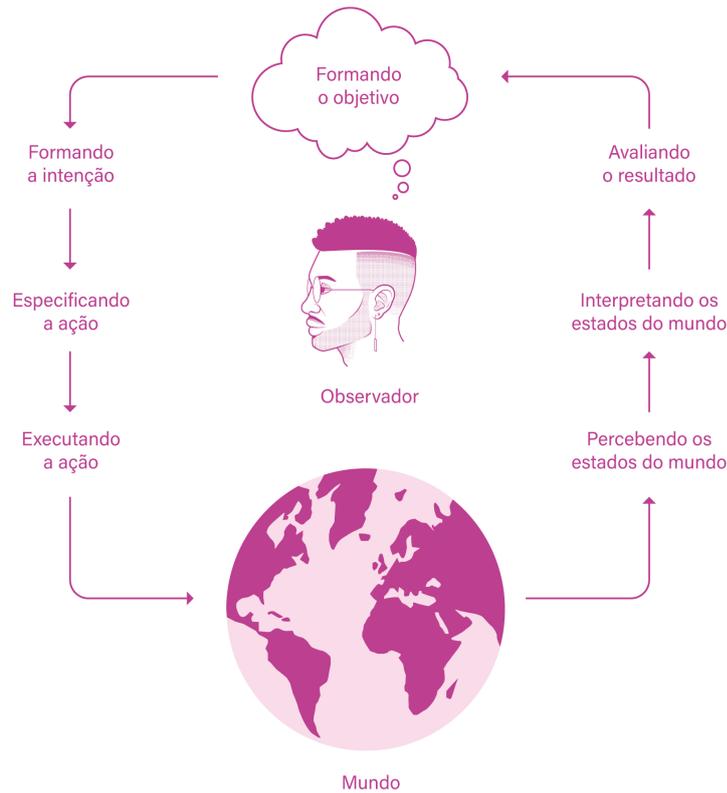


Figura 09 - Sete Estágios de Ação (Fonte: Dubberly, Pangaro e Haque, 2009, p. 2)

Veloso (2011), assim como Dubberly et al. (2009), retoma em sua pesquisa o modelo de Verplank (2001), explicando que o modelo também é baseado no ciclo de feedback - chamado-o de *How do you...feel-know-do* – em seu livro Verplank explica que é preciso ter em mente três perguntas: como o usuário se sente (*feel*) - esse momento é relacionado ao *feedback*; como ele sabe (*know*), quais caminhos ele trilha para interagir com o sistema; e o como ele faz (*do*) isso, que é relacionado à maneira pela qual o usuário afeta o mundo (Veloso, 2011, p 13).

O modelo de Verplank (2001, *apud* Dubberly et al., 2009, p. 3), é uma relação entre mãos e botões como podemos observar na Figura 10.

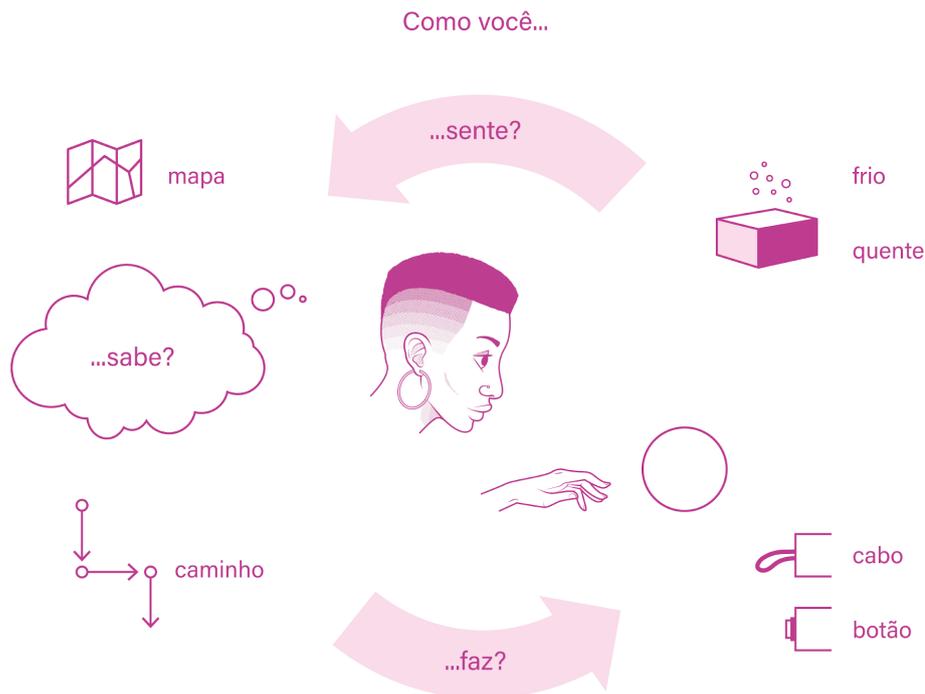


Figura 10 - Modelo de como sente-conhece-faz (Fonte: Dubberly, Pangaro e Haque, 2009, p. 3)

Dubberly *et al.* (2009, p. 3) acreditam que o sentir e o agir preenchem a lacuna entre usuário e sistema. Por isso, representar a interação entre humanos e sistemas com o modelo de *feedback-loop* é o ponto inicial para entender o que é interação humano-máquina.

Até o momento, nesta pesquisa observamos diversas definições e olhamos para alguns modelos construídos que exemplificam a interação humano-máquina, ressaltando, por fim, a importância do modelo *feedback-loop*.

O modelo proposto por Kroemer e Grandjean (2007, p. 125) retrata como a ergonomia humano-máquina era tratada antes mesmo da era digital, na qual vivemos hoje. Partindo de máquinas decorrentes da revolução industrial do século XIII, estudava-se a relação do humano (operário) com a máquina na qual ele realizava suas tarefas. Na Figura 11 podemos observar o esquema proposto por Kroemer e Grandjean:

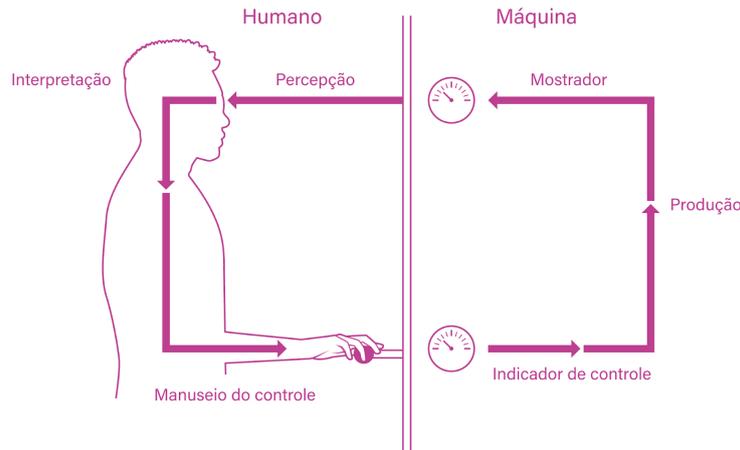


Figura 11 - Ergonomia humano-máquina (Fonte: Kroemer Grandjean, 2007, p. 125)

Kroemer e Grandjean (2007, p. 125) afirmam que através da imagem é possível, além de representar a dinâmica homem-máquina, inferir que este modelo de sistema – também denominado por ele como um sistema fechado – aponta para o ser humano como peça importante na construção do conjunto, já que, "compete a ele o poder de decidir".

Os autores descrevem todo o processo mostrado na Figura 11 iniciando pelo mostrador, que é a fonte para a tomada de decisões, já que é através dele que o sistema entrega informações sobre o que está acontecendo à pessoa que irá operá-lo. Por meio do mostrador, o humano recebe as informações de forma visual e, através da sua percepção, juntamente com seus conhecimentos obtidos de interações anteriores, interpreta as mensagens e então é capaz de agir. A seguir, partindo do momento ao qual toma suas decisões com a intenção de agir, o humano deve manusear corretamente os controles, para que então a máquina consiga informar quais foram os resultados deste *input* e seguir para o próximo passo, realizando o processo para o qual foi programada, fechando assim o ciclo no momento em que a máquina, através de seus mostradores, indica que os acontecimentos característicos da interação foram de fato realizados.

Kroemer e Grandjean (2007, p. 126) afirmam que, em momentos passados, as máquinas não eram difíceis de serem compreendidas e de serem manuseadas, mas que a cada conquista tecnológica, aperfeiçoam-se os controles sobre elas e o capitalismo avança. Dessa maneira, a relação homem-máquina torna-se cada vez mais complexa, o que sem dúvidas tem como principal influenciador o próprio capitalismo, que tende a aumentar tanto o nível de consumo como o de produção.

Por outro lado, o homem em harmonia com a máquina é capaz de evitar grandes catástrofes, como por exemplo, a que se relaciona ao impedimento de eventos inesperados numa usina nuclear.

As definições dadas por Barbosa e Silva (2010) e por Dubberly e Pangaro (2019), quanto à apresentada em seus primórdios por Kroemer e Grandjean (2007), parecem convergir para um lugar comum.

3.2.2

O papel das atividades mentais nas interações

Neste tópico, abordaremos de forma sucinta como as atividades mentais estão relacionadas à interação entre o humano e as máquinas e quais são seus papéis nesse processo.

Kroemer e Grandjean (2007, p. 141) dividem as atividades mentais em dois grupos: atividades mentais no sentido *lato* e processamento de informações como parte do sistema "homem-máquina". O primeiro grupo é referente ao uso da criatividade e o quanto dela é exigido. De forma simples, a criatividade é o cruzamento de informações com os conhecimentos técnicos, de maneira a produzir novas ideias. Fatores como conhecimento, experiência, agilidade cognitiva e habilidade de criação são decisivos para a utilização desse grupo dentro das atividades mentais. Já o segundo grupo é relacionado ao processamento de informações do que foi discutido durante toda a seção anterior deste trabalho, quando exploramos os sentidos de interação. Contudo, sobre esse assunto, Kroemer e Grandjean (2007, p. 141) ressaltam os três pontos fundamentais desse grupo, os quais são: percepção, interpretação e processamento da informação transmitida pelos órgãos dos sentidos. Os dois primeiros pontos são bem delineados pelo autor quando o mesmo descreve o processo de interação fechado humano-máquina - descrito anteriormente. Resta trazer aqui o que ele define como processamento da informação transmitida pelos órgãos dos sentidos.

Explicitando o processamento da informação transmitida pelos órgãos dos sentidos, podemos dizer que o "'processamento' consiste na combinação de nova informação com o que já é sabido, fornecendo a base para a tomada de decisão." (Kroemer e Grandjean, 2007, p. 141).

Em relação à carga mental, o autor define que certas condições do ambiente de trabalho são determinantes. Ele lista quatro condições:

1. A obrigação de manter um nível elevado de alerta durante longos períodos.
2. A necessidade de tomar decisões que envolvem grande responsabilidade para a qualidade do produto e para a segurança de pessoas e equipamentos.
3. Uma redução ocasional da concentração devido à monotonia.
4. Falta de contato humano quando um posto de trabalho é isolado de outros. (Kroemer e Grandjean, 2007, p. 142)

Vale ressaltar que essas definições se enquadram no contexto de trabalho fabril, no qual operários lidam com máquinas a fim de cumprir suas funções. Os autores pontuam que áreas como a neurofisiologia e a psicologia, dentre outras, têm mostrado grande esforço para elaborar uma construção em torno do que ele chama de "processos básicos de esforço mental" (Kroemer e Grandjean, 2007, p. 142). Ainda segundo esses autores, dentre os resultados trazidos das áreas ligadas aos estudos de processos mentais, existem três que nos interessa em torno da ergonomia, que são: a captação de informações, a memória e a manutenção do estado de alerta.

3.2.2.1 Captação de informações

A captação de informações está ligada à teoria de Shannon e Weaver que, através de um modelo matemático, desenvolveram o que conhecemos como *bit*, que é o que define uma unidade de informação (Kroemer e Grandjean, 2007, p. 142). Para os autores "um bit é uma unidade de informação que consiste de uma alternativa de duas opções possíveis de mesma natureza". Contudo, com os aumentos do número de alternativas, essa relação se torna mais complicada. Para ilustrar, imaginemos um ser humano dirigindo um carro. Nesse contexto, ele recebe *N* informações que não são possíveis de serem medidas, pois não se trata apenas de sinais unitários codificados, o que aumenta a complexidade na relação humano-máquina.

Outra teoria abordada por Kroemer e Grandjean (2007, p. 142) é a teoria da capacidade de canal, a qual afirma que a entrada de informações está ligada aos órgãos dos sentidos humanos. Segundo a teoria, as informações são recebidas pelos canais, mas não há como medi-las, a não ser quando há saída de informações por

estes canais, o que torna possível mensurar a capacidade do canal de acordo com as informações recebidas.

Ainda segundo essa teoria, caso a quantidade de informações recebidas por determinado canal seja pequena, quando for necessário que haja a saída desta informação, ela não será reduzida, mas caso a massa de informações seja muito grande, será percebido um *déficit* na quantidade de informações que foram captadas. Esse limite na saída de informações é o que define a chamada "capacidade de canal".

Através de experimentos ligados aos canais visuais e auditivos, foi definido que a maior capacidade do canal está diretamente ligada à recepção da fala. Em termos comparativos, temos o humano que é capaz de receber um vocabulário de 2.500 palavras, algo entre 34 e 42 *bits* por segundo e um telefone que tem a capacidade de 50.000 *bits* por segundo (Kroemer e Grandjean, 2007). Observando que no dia a dia do ser humano há muito mais recebimento de informações do que saída. Kroemer e Grandjean (2007, p. 143) afirmam que "na vida diária, a informação é maior do que a capacidade de canal do sistema nervoso central, de forma que tem que ocorrer uma considerável 'redução de processamento!'".

3.2.2.2

Memória

A memória é um grande armazenamento de informações recebidas. É por meio dela que se recorre aos nomes de amigos, datas importantes e festivas, acontecimentos da infância e outras lembranças que são capazes de marcar a trajetória dos seres humanos. Segundo Kroemer e Grandjean (2007, p. 143), é através desses momentos marcantes que conseguimos definir o que é relevante ou não para ser armazenado. Para isso, podemos distinguir dois tipos distintos de memória: a memória de curta duração (recente) e a de longa duração.

A memória de curta duração está diretamente relacionada a acontecimentos que ocorreram nos últimos minutos e horas; a de longa duração relaciona-se aos acontecimentos dos últimos meses e anos.

Kroemer e Grandjean explicam que:

Um modelo para a memória de curta duração considera que a informação recebida deixa traços que continuam a circular como um estímulo dentro da rede neuronal, sendo que, por um tipo de retroalimentação (*feedback*), pode ser recuperada para a esfera consciente a qualquer momento (Kroemer e Grandjean, 2007, p.143)

Já para armazenar algo na memória de longa duração é possível supor que existe um período de tempo ao qual é necessário para a gravação dessa memória em nosso cérebro. Durante esse período de tempo, alguns rastros destas memórias são apagados ou danificados, resultando posteriormente em engramas¹³ muito mais resistentes, sendo, então, armazenados na memória de longa duração, segundo Kroemer e Grandjean (2007).

Os autores chamam a nossa atenção para uma parte específica do cérebro: o sistema límbico, com destaque aqui para o hipocampo, que para esses autores, é uma das partes do cérebro responsável pela gravação da memória de curta duração. Eles afirmam que o sistema límbico é também um dos responsáveis por boa parte das emoções, como raiva e medo, além de ser ligado ao comportamento sexual e à motivação do ser humano. Junto ao sistema límbico há o que os autores denominam de sistema de ativação da formação reticular, que desempenha um papel importante, uma vez que está diretamente ligado ao processo de aprendizado e à memória. Kroemer e Grandjean atribuem a esse processo o chamado estado de vigilância, que é o verdadeiro responsável pelo armazenamento da memória:

[...] A formação reticular é um sistema de ativação essencial para os processos de aprendizado e de memória, já que apenas durante o estado de vigília pode ocorrer o armazenamento da memória. Quanto mais ativo o cérebro, mais aprendizado e memória ocorre, e, sem dúvida, mais alto é o estado de motivação. (Kroemer e Grandjean, 2007, p. 144)

3.2.2.3

Atenção prolongada (vigilância)

Kroemer e Grandjean (2007) ressaltam a importância das características ligadas às capacidades mentais que são relacionadas nesta seção, que são os seguintes pontos: A) os tempos de reação e B) a capacidade mental, a variabilidade

¹³ "Traço permanente deixado na psique por tudo que tenha sido experimentado psiquicamente; traço latente de memória." (Michaelis, 2020)

da frequência cardíaca, dentre outros. Abordaremos a seguir, de forma sucinta, os tópicos que interessam a esta pesquisa, a fim de elucidar sua importância para as atividades mentais humanas:

A) Tempo de reação

O tempo de reação humana é estudado tanto pela psicologia quanto pela ergonomia, com distintas intenções. O estudo do tema se mostra importante para a psicologia, porque através dele é possível observar os processos mentais. Por outro lado, para a ergonomia, o tempo de reação pode ser utilizado como parâmetro para realização de atividades mentais (Kroemer e Grandjean, 2007).

Os autores mostram, segundo Wargo (1967 apud Kroemer e Grandjean, 2007), algumas durações de tempo relacionadas ao humano e suas reações: mudanças nos órgãos dos sentidos, de 1 a 38 ms; transmissão neural até o córtex cerebral, de 1 a 100 ms; elaboração central dos sinais, de 7 a 300 ms; transmissão neural para os músculos, de 10 a 20 ms; e, tempo de latência do músculo, de 30 a 70 ms. Os autores afirmam que o tempo de reação total não pode ser menor do que 100 ms, já que há também o tempo adicional do movimento que sucede a reação. Os autores ainda exploram alguns outros conceitos que se aprofundam na questão tempo de reação, como por exemplo, o tempo de reação simples, e o tempo de reação seletiva.

B) Limites de carga mental

Todos nós sabemos que é impossível manter a atenção em uma atividade por um longo período de tempo. Imagine que estamos em uma palestra superinteressante e prestando atenção nela desde que foi iniciada. Após quatro horas, nossa atenção ou será diminuída, ou focará em outra coisa. Os próprios autores ilustram esse fato com outra situação: quanto mais lemos um mesmo material por horas, mais precisaremos lê-lo novamente, para que possamos compreendê-lo.

Kroemer e Grandjean (2007) destacam o trabalho de outros pesquisadores, especialmente da área de psicologia, como Bills (1931) e Broadbent (1958), acerca da dificuldade de concentração em uma tarefa após muitas horas a ela dedicadas, sem oportunizar momentos de pausa. Os bloqueios encontrados por Bills (1931 apud Kroemer e Grandjean, 2007, p. 147) foram definidos pelos autores como "[...] interrupções em intervalos frequentes, no processamento de informação [...]", acabam acontecendo, como demonstram os exemplos dados anteriormente, gerando

fadiga, já que quanto maior for o tempo de trabalho mental, mais e maiores interrupções acontecem. Por isso, os autores também destacam as descobertas de Broadbent (1958) as quais indicam que o aparecimento de sobrecargas - como ruídos e cansaço - contribui para o aumento dos bloqueios, já que o indivíduo está em uma mesma tarefa mental por muito tempo, fatigando-se mentalmente. (Kroemer e Grandjean, 1998, p. 147)

Os pontos (A) e (B) apresentados acima deixam evidente a importância da atenção prolongada, quando se trata de recepção de informações pelos seres humanos. Percebemos aqui a existência de limites para a recepção de informações em uma interação humano-máquina e que esse limite está diretamente relacionado ao tempo dedicado ao esforço mental e também no armazenamento da memória.

Quando falamos de interação humano-computador, e ainda mais quando evidenciamos o recorte - como nesta pesquisa - na interação através do som não podemos esquecer de que o ser humano é um dos sistemas que integram o ciclo de interação. Entender os sistemas participantes é fundamental para que elaboremos aqui, nesta pesquisa, futuros desdobramentos.

4

Interação sonora

Quando focamos na questão de interação sonora entre o humano e o computador, com o objetivo de compreender a extensão numa conversação, devemos atentar para o fato de que uma das pontas desta equação é uma máquina e que essa não conversa e interage como seres humanos.

Para continuarmos a discussão sobre o objeto desta pesquisa é necessário caracterizar como esta interação sonora se apresenta – em qualquer uma das duas pontas deste fio comunicador que é a conversa.

Para tal, este capítulo aborda, primariamente, o conceito de cognição, peça fundamental para a compreensão dos estímulos que recebemos do ambiente que nos cerca e que os transforma em conhecimento. Posteriormente, abordamos o conceito de interação por voz.

4.1

Cognição Auditiva

A cognição – apesar de ser praticamente imperceptível no nosso dia a dia – é fundamental para a compreensão do que nos cerca e para o aprimoramento das nossas capacidades humanas. Nas palavras de Neuhoff (2011) podemos perceber a sutileza do processo cognitivo:

A percepção é quase sempre um processo automático e fácil. A luz e o som no ambiente parecem, quase magicamente, transformados em uma complexa gama de impulsos neurais que são interpretados pelo cérebro como uma experiência subjetiva das cenas auditivas e visuais que nos cercam. Essa transformação de energia física em "significado" é concluída em uma fração de segundo. (Neuhoff, 2011, p. 63¹⁴, tradução nossa)

¹⁴ "Perception is almost always an automatic and effortless process. Light and sound in the environment seem to be almost magically transformed into a complex array of neural impulses that are interpreted by the brain as the subjective experience of the auditory and visual scenes that surround us. This transformation of physical energy into "meaning" is completed within a fraction of a second".

Cada uma dessas pequenas experiências soma-se às nossas memórias e com isso vamos criando um banco de dados particular. Ao olhar para o céu quando está branco de nuvens, podemos supor que está nublado, que pode chover e que provavelmente está frio. Podemos reconhecer prontamente, na ponta da língua, o gosto da pimenta e sabemos o ardor que está por vir. Ao vermos a fumaça saindo da água na chaleira, entendemos que aquele líquido está quente.

São tantos *inputs* que recebemos a todo instante, que nem nos damos conta. A mesma coisa acontece com os sons. Sabemos quando estamos nos aproximando da cachoeira muito antes de vê-la; sabemos que há um curto-circuito, mesmo que não vejamos as fagulhas; sabemos que é um ente querido ao telefone, mesmo que não possuamos um identificador de chamadas.

A cognição sonora acontece tanto quanto as cognições ligadas aos nossos outros sentidos. Entretanto, não há um largo estudo neste campo de pesquisa, conforme nos aponta Neuhoff (2011):

Há uma rica história da pesquisa psicoacústica sobre os aspectos "sensoriais" da audição. Por outro lado, em termos comparativos, a "cognição auditiva" recebeu pouca atenção. As informações acústicas recebidas são transformadas em um sinal neural no nível das células especializadas, no ouvido interno. Com exceção da fala e da música, é aqui que as pesquisas no campo da audição frequentemente estagnam-se. (Neuhoff, 2011, p. 77¹⁵, tradução nossa)

O *input* sonoro que ouvimos não vem vazio de significado. Assim que captado, ele se mescla com memórias, conhecimento e capacidades pessoais de cada indivíduo, ganhando significado a partir disso, sendo através dessa mescla que o indivíduo pode gerar percepções acústicas acerca da informação fornecida. (Neuhoff, 2011)

Atualmente, há o entendimento da importância desta área e as pesquisas sobre o tema começam a ser desenvolvidas. Entretanto, por ser um campo novo e extremamente complexo, ainda há pouquíssima compreensão sobre isso, conforme o autor explicita, quando descreve que estudos comportamentais e de neuroimagem

¹⁵ *"There is a rich history of psychoacoustic research on the "sensory" aspects of audition. Conversely, "auditory cognition" has received comparatively little attention. Incoming acoustic information is transformed into a neural signal at the level of specialized cells in the inner ear. With the exception of speech and music, this is where the study of audition often stopped."*

forneem evidências de forma experimental de que o cérebro armazena representações de estímulos do mundo externo.

Os estudos nos mostram que, assim como a visão, o som pode gerar memórias. A partir daí, a sonoridade pode ser usada como ferramenta cognitiva, tanto quanto as imagens são usadas. Assim, Neuhoff (2011) conclui que:

Há, também, evidências de representações cognitivas de estímulos acústicos no domínio auditivo. Como no domínio visual, evidências subjetivas e informais são abundantes. Quase todo mundo admite ser capaz de imaginar o som de uma buzina de carro, do canto de pássaros ou de ovos fritando em uma frigideira. (Neuhoff, 2011, p. 78¹⁶, tradução nossa)

Para tentar dar conta do potencial existente no tema do campo sonoro, temos uma recente área de estudo que se denomina *Sonic Interaction Design* - ou *Design de Interação Sonora (DIS)*. Essa área insere o tema sob a ótica do estado da arte do *Design* e, assim, podemos apreciar o tema por diversos focos diferentes, em busca da melhor utilização do som. Serafin, Franinovic, Hermann, Lemaitre, Rinott e Rocchesso (2011) nos traz que:

O *Design de Interação Sonora (DIS)* é uma área de estudo interdisciplinar que surgiu recentemente devido à dedicação conjunta de pesquisadores e de profissionais que trabalham na interseção da computação sonora e musical, do *design* de interação, da interação humano-computador, das novas interfaces para expressão musical, do *design* de produtos, da psicologia e da cognição musical, da composição musical, da performance e das artes interativas. (Serafin *et al.*, 2011, p. 87¹⁷, tradução nossa).

A partir daí podemos vislumbrar um estudo que mantenha sob sua ótica o som e correlacione esse ao usuário, ao serviço, ao produto e ao ambiente em que está inserido, e não mais o som pelo som. Conforme apontam os autores, quando descrevem que na computação, seja ela sonora ou musical, pesquisadores se afastaram da simples reprodução dos sons musicais e do cotidiano humano, abrindo

¹⁶ "In the auditory domain, there is also evidence for cognitive representations of acoustic stimuli. As in the visual domain, there is abundant subjective and anecdotal evidence. Almost anyone will admit to being able to imagine the sound of a car horn, a bird chirping, or of eggs frying in a pan."

¹⁷ "Sonic Interaction Design (SID) is an interdisciplinary field which has recently emerged as a combined effort of researchers and practitioners working at the intersection of sound and music computing, interaction design, human-computer interaction, novel interfaces for musical expression, product design, music psychology and cognition, music composition, performance and interactive arts."

espaço para investigações capazes de auxiliar no design de som para além do que existia.

Como o som está ligado à cognição, ele faz com que o receptor perceba o mundo que o cerca, capacitando-o a entender o momento em que está ou compreender o conhecimento com o qual está lidando. Serafin *et al.* (2011) propõe que:

O som em interação é, certamente, um fenômeno multifacetado que pode ser compreendido em diversos níveis, como o nível estético, emocional, afetivo, de coordenação, de informação e até mesmo o nível social e cultural. (Serafin, *et al.*, 2011, p. 105¹⁸, tradução nossa).

Tendo essa percepção sobre o som, há todo um novo campo a ser explorado em que a sonoridade não é uma peça qualquer, mas sim um agente fundamental na relação na qual se insira. Os autores salientam que:

O desenvolvimento do DIS segue as tendências da chamada terceira onda de interação humano-computador, na qual a cultura, a emoção e a experiência, no lugar da função e da eficiência apenas, estão inclusas na interação entre humanos e máquinas. (Serafin *et al.*, 2011, p. 107¹⁹, tradução nossa)

Assim, lidar com o som, entendendo-o como ferramenta cognitiva, amplia nossas percepções sobre a importância do tema, abrindo-se também as possibilidades de estudo.

4.2

Interação por voz

A interação por voz não é nada mais do que uma conversa, mesmo que esta conversa – ou interação – seja entre ser humano e máquina. E sendo a conversa uma das diversas formas de comunicação, precisamos utilizar alguma linguagem para que consigamos obter êxito nesta atividade.

¹⁸ “*Sound in interaction is certainly a multi-faceted phenomenon which can be understood on various levels including the aesthetic, emotional, affective, coordination, information and even social and cultural level.*”

¹⁹ “*The development of SID follows the trends of the so-called third wave of human-computer interaction, where culture, emotion and experience, rather than solely function and efficiency, are included in the interaction between humans and machines.*”

Além dos Ícones Auditivos e dos Ícones Sonoros, há uma forma de interação sonora conhecida como *Spearcon*, na qual é possível ir além dos sons do mundo e da música abstrata e se chega a uma fala propriamente dita. McGookin e Brewster (2011) mostram que:

Spearcon é a fala sintetizada que é acelerada na medida em que não é mais ouvida como fala. Essa impressão digital auditiva exclusiva pode ser usada para representar itens em uma estrutura de menu e, nos testes iniciais, mostrou que um desempenho mais rápido pode ser obtido em relação aos Ícones Auditivos e aos Ícones Sonoros, com uma quantidade significativamente menor de treinamento necessário. (McGookin; Brewster, 2011 p. 351²⁰, tradução nossa)

Dessa maneira, compreendemos que estamos somente no começo da estrada para podermos ter uma comunicação eficaz com nossos computadores. Mas também compreendemos que já possuímos as ferramentas necessárias para desenvolvimento nesta área. Sejam elas instrumentos computacionais, como o Ícones Auditivos e os Ícones Sonoros, sejam elas instrumentos da nossa comunicação humana – linguagem, signos, cognição. O desafio agora é imbuir nossas máquinas com ferramentas que sejam capazes de dar conta tanto da nossa compreensão sobre elas, quanto da capacidade delas de nos compreender. Afinal, uma conversa precisa ter um fluxo sempre em movimento.

Hall (2018) introduz o pensamento de que a forma com a qual seres humanos interagem entre si deveria ser a base de toda e qualquer interação com sistemas digitais, já que para nós interagir através da fala é a forma mais fácil e divertida e, por vezes, até mesmo dolorosa.

Segundo a arqueóloga Uomini e o psicólogo Meyer, nossos ancestrais começaram a desenvolver a linguagem em torno de 1.75 milhões de anos atrás (Hall, 2018), como podemos observar na Figura 12, a seguir:

²⁰ "*Spearcons, synthesised speech that is speeded up to the extent that it is no longer heard as speech. This unique auditory fingerprint can be used to represent items in a menu structure and in initial tests has shown that faster performance can be obtained over both Auditory Icons and Earcons, with significantly less training required.*"

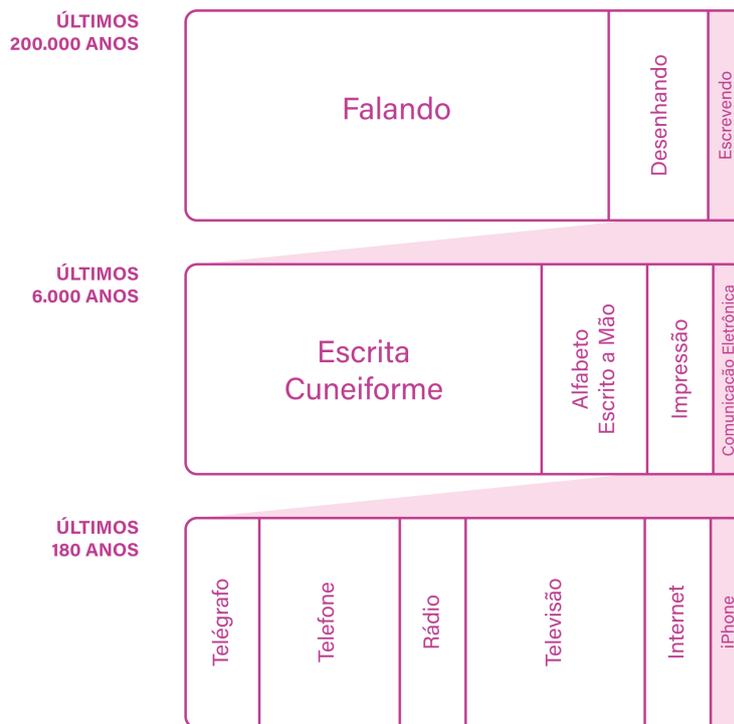


Figura 12 - Evolução das linguagens (Fonte: Hall, 2018, p. 11, reprodução nossa)

Mesmo com toda a evolução que mudou o mundo e a forma com a qual os seres humanos se comunicam através dos séculos, Hall (2018, p. 12) afirma que tais alterações não mudaram o nosso núcleo central, a oralidade.

Como vimos no segundo capítulo dessa pesquisa, Martino (2005) aponta que o cerne comunicacional da comunidade primitiva era a fala. Para Hall (2018), essa forma de linguagem foi o primeiro modelo de interface.

Antes da alfabetização, as palavras eram efêmeras e todo o conhecimento era social e comunitário. Não havia opção de "salvar" nem propriedade intelectual. A única maneira de sustentar uma ideia era compartilhá-la, falando em voz alta com outra pessoa de uma maneira que a tornasse realidade fácil para eles lembrarem. Essa foi a oralidade - a primeira interface. (Hall, 2018, p. 44)

Ora, se a fala é, de acordo com Hall (2018), o primeiro modelo de interface, por que só agora, na segunda década dos anos 2000, estamos voltando o olhar para interfaces que permitem a interação por meio da voz?

Na realidade, estudos na área da interação por voz não são tão recentes, como já vimos anteriormente. Pearl (2016), assim como Warren (2014), mostram-nos que pesquisas na área são realizadas desde, pelo menos, os anos de 1950, quando o *Bell Labs* construiu o primeiro sistema de reconhecimento de dígitos. Esse sistema que, até então tinha seu vocabulário restrito e não possuía muita utilidade fora do laboratório, evoluiu. Nas décadas seguintes (1960 e 1970), com a continuação das pesquisas, o número de palavras compreendidas pelo sistema aumentou e também direcionou para o que compreendemos como reconhecimento de fala contínuo, ou seja, esses estudos tornaram possíveis à compreensão da fala, sem a necessidade de pausas elaboradas após cada palavra. (Pearl, 2016)

Os estudos da área não pararam e em 1980 os avanços permitiram que o reconhecimento de fala contínua fosse firmado como uma realidade para a próxima década, que veio a ser conhecida como a primeira grande era da interface por voz do usuário. A autora ainda acrescenta que, em 1990, tornou-se viável o primeiro alto falante independente, o que significa que qualquer ser humano poderia falar com o sistema. (Pearl, 2016)

Segundo Pearl (2016), a primeira grande era da interface por voz do usuário se deu principalmente pelos sistemas de resposta de voz interativa (RVI)²¹, que eram capazes de compreender a fala humana por meio do telefone.

Qualquer pessoa que, por exemplo, tenha vivido dos anos 2000 até os tempos atuais e tenha uma conta de telefone, a qual precisou cancelar, esteve em contato direto com este tipo de sistema. Durante a primeira década dos anos 2000, sistemas RVI se tornaram populares já que, segundo Pearl (2016), qualquer pessoa com um telefone conseguia obter cotações de ações, realizar reservas em voos, transferir dinheiro entre contas, informar-se sobre sessões de cinema e até mesmo pedir informações sobre o tráfego. Tudo isso através da voz e de um telefone fixo.

Por mais que esse tipo de sistema tenha obtido uma má reputação ao longo do tempo, rendendo inclusive esquetes em programas como o *Saturday Night Live*²², eles também foram o que Pearl (2016, p. 2) chama de "boom", como é possível perceber nesta passagem: "[...] no início dos anos 2000, uma empresa de frete recebeu inúmeras ligações furiosas depois que seu sistema RVI foi retirado

²¹ *Interactive voice response (IVR)*

²² Programa de humor norte americano.

para manutenção [...]". Podemos afirmar, portanto, que a primeira era da interface por voz do usuário foi marcada pelo desenvolvimento e democratização dos sistemas RVIs.

4.2.1

O poder das interações na segunda era das interfaces por voz

Pearl (2016) divide a história das interfaces de voz em dois grandes momentos: a primeira era das interfaces por voz, explicitado acima, e o segundo, denominado como "segunda era", a qual abordaremos neste tópico.

A autora caracteriza a segunda era principalmente pelo surgimento de sistemas como a *Siri* (Apple), *Google Now* (Google), *Hound* (SoundHound) e *Cortana* (Microsoft), que combinam informações visuais e auditivas nas interações com seus usuários. Mais recentemente, temos o que pode ser considerada a evolução do *Google Now*, o *Google Assistente*, e outros sistemas que funcionam exclusivamente através da voz, como o *Amazon Echo* (Amazon) e o *Google Home* (Google).

Dunn (2016 apud Pearl, 2016) aponta o crescente aumento na utilização desses sistemas quando cita dos obtidos pela empresa de inteligência de mercado Tractica²³ que afirma que até o ano de 2021 o uso de assistentes virtuais passará de 504 milhões para 1,8 bilhão, ou seja, estamos vivendo um período importante para os assistentes virtuais e para os modelos de interfaces que abrangem esse universo, o que destaca a relevância dos estudos na área.

Dunn (2016 apud Pearl, 2016) ressalta também que os sistemas como *Siri* (interface de voz da Apple) e *Alexa* (interface por voz do *Amazon Echo*) estão em expansão para além de *smartphones* e alto-falantes, através da internet das coisas, e ressalta o desejo de que conversas com sistemas se tornem mais naturais.

Outra informação importante é destacada por Pearl (2016), quando indica que a *Google* reportou que pelo menos 20% das pesquisas feitas hoje acontecem via voz. Ela ainda afirma que estamos vivendo hoje o que pode ser compreendido como a infância dessa era, já que há muitas coisas que nossos *smartphones* e dispositivos podem realizar, mas também existem muitas outras que ainda não.

²³ Empresa de inteligência de mercado que se concentra em tecnologias emergentes.

É nesse cenário que ressaltamos, por intermédio de Pearl (2016) e Hall (2018), as qualidades da oralidade e das interfaces por voz.

Iniciamos por Hall (2018), que defende a linguagem oral e a ideia advinda de Ong (1982), quando explora a psicodinâmica²⁴ da oralidade, dizendo que oralidade e alfabetização coincidem. Hall (2018, p. 14) aponta que, Ong (1982) em sua pesquisa revelou que, ao preservarmos nossas ideias através da escrita, aumentamos nossos conhecimentos, assim como alteramos nossos valores e comportamentos. Ele destaca que pessoas que não tiveram contato algum com a escrita são diferentes daquelas que tiveram. Hall (2018, p. 14) ainda aponta - assim como em Martino (2005) - que essas pessoas dependem uma das outras para compartilhar e perpetuar seus conhecimentos, o que cria uma relação diferente e mais íntima entre a ideia e a comunidade.

Hall (2018, p. 15) acredita que a cultura oral é imediata e social quando define que em uma sociedade, a qual não se apoia na escrita, a forma de se comunicar acontece apenas no momento e "cara a cara". Além disso, explica que as palavras, quando ditas, se tornam eventos que existem apenas naquele momento, sendo impossível recuperar o que foi dito ou examinar o que foi falado, por mais que a fala possa ser repetida pelo orador, já que não há como registrar ou reproduzir expressões relacionadas aquela fala.

Hall (2018) também afirma que todo conhecimento é social e vive apenas na memória; que as pessoas precisam estar presentes para que haja a troca de conhecimento ou comunicação, sendo que toda comunicação é participatória e imediata, tendo o orador o poder de ajustar seu discurso ao contexto; que nas sociedades orais, quem detém o conhecimento, é a própria comunidade e não os indivíduos, sendo impensável a reivindicação de uma ideia por parte de alguém; e, por fim, que não há padrão no uso das linguagens orais em uma comunicação, já que não existem dicionários ou instituições autorizadas a determinar como uma palavra deva ser dita.

Pearl (2016), por outro ângulo, apresenta especificamente as qualidades da voz em quatro aspectos, sendo esses: velocidade, mãos livres, intuitividade e empatia, que podemos observar, conforme se segue:

²⁴ Conjunto de fatores de natureza mental e emocional que motivam o comportamento humano, especialmente os que aparecem como reação inconsciente aos estímulos ambientais.

- Velocidade: um estudo da universidade de Stanford mostra que ditar mensagens escritas é mais rápido do que digitá-las²⁵;
- Mãos livres: em casos nos quais usuários estão cozinhando ou dirigindo, a fala se torna muito mais prática, se comparada à digitação;
- Intuitividade: todo mundo sabe falar;
- Empatia: é difícil para os seres humanos entenderem o tom por meio da escrita, a exemplos de *e-mails* mal interpretados. (Pearl, 2016, paginação irregular)

Além de ressaltar os aspectos positivos no uso da interface de voz, Pearl (2016), também segrega em dois grupos os tipos de dispositivos (elencados na introdução desta pesquisa) capazes de realizar interações através da voz. A autora faz uma divisão nos modelos de dispositivos, agrupando assistentes como *Amazon Echo* e *Google Home* no que ela denomina como “assistentes de casa”²⁶, destacando-os do que conhecemos como assistentes de celulares²⁷ (que nesta pesquisa chamamos de assistente virtual inteligente). Existem outros modelos de dispositivos, como os *wearables*, TVs, carros e até mesmo brinquedos, conforme destaca Pearl (2016). No entanto, para recorte de estudo desta pesquisa, vale ressaltar a importância dos assistentes virtuais inteligentes e os definidos como assistentes de casa.

4.3

Conversa como elemento primordial da interação

A interface fundamental entre as pessoas é a conversa (Hall, 2018, p. 25, tradução nossa²⁸).

Ao pensarmos em interação, uma das principais formas na qual a relacionamos é com a conversa. É com ela, geralmente empregada a partir da oralidade, que podemos fazer a troca do diálogo. Nesse diálogo podemos ceder parte do que é nosso e receber algo do outro. É com esse diálogo que alcançamos novas coisas a partir do que já possuíamos em conjunto com o que ganhamos da outra ponta deste sistema. Por meio do diálogo nos perpetuamos e evoluímos.

²⁵ Shahani, A. (2016). “*Voice Recognition Software Finally Beats Humans At Typing, Study Finds*”. Disponível em <http://npr.org/>. Acesso em: 11 jul. 2020.

²⁶ “*Home Assistants*”.

²⁷ “*Mobile Phone Assistants*”.

²⁸ *The fundamental interface between people is conversation.*

Para entender o cerne da conversação, podemos segmentá-la a partir desta análise feita por Pangaro (2017, tradução nossa²⁹):

- # contexto - um momento, situação, lugar e/ou história compartilhada que prepara o cenário para [...];
- # linguagem - ao menos um significado inicial compartilhado que transmita sentido, para começar uma [...];
- # troca - engajamento recíproco, em interações baseadas em linguagens que podem resultar em [...];
- # acordo - (suficientemente) entendimento compartilhado de conceitos, intenções, valores, que podem levar a [...];
- # ação ou (trans)ação - uma interação coordenada em domínios que não sejam a linguagem, por exemplo, comércio, contratos, dança, jogos. (Pangaro, 2017, p. 1580, tradução nossa)

A conversa, essa complexa interação, deve ser advinda de um objetivo, como Pangaro (2017, p. 1580, tradução nossa)³⁰ destaca: "Conversas se iniciam com um participante tendo algum tipo de objetivo, seja específico ou geral, articulado ou ainda não formado". Sem algum desses itens a interação não ocorre ou ocorre de forma comprometida. Sem um objetivo não há por que a conversa se iniciar; sem um contexto em comum as partes, podem até nunca se encontrar. A linguagem é fundamental para o entendimento, pois sem a troca, passamos a ter um monólogo e não um diálogo. Sem entendimento passamos a ter uma conversa improdutiva, que não tem poder evolutivo. E, temos por fim, a ação que é a única que pode fazer ou não parte da conversa, não alterando sua integridade, como podemos observar na figura a seguir:

²⁹ # context – a moment, situation, place and/or shared history sets the stage for [. . .];
 # language – at least an initial shared means for conveying meaning, to begin an [. . .];
 # exchange – engagement in back-and-forth, language-based interactions that may build to [. . .];
 # agreement – (sufficiently) shared understanding of concepts, intent, values, that may lead to [. . .]; and
 # action or (trans)action – a coordinated interaction in domains other than language, for example, commerce, contracts, dance, games.

³⁰ "Conversations begin with a participant having some sort of goal, whether specific or general, articulated or unformed."

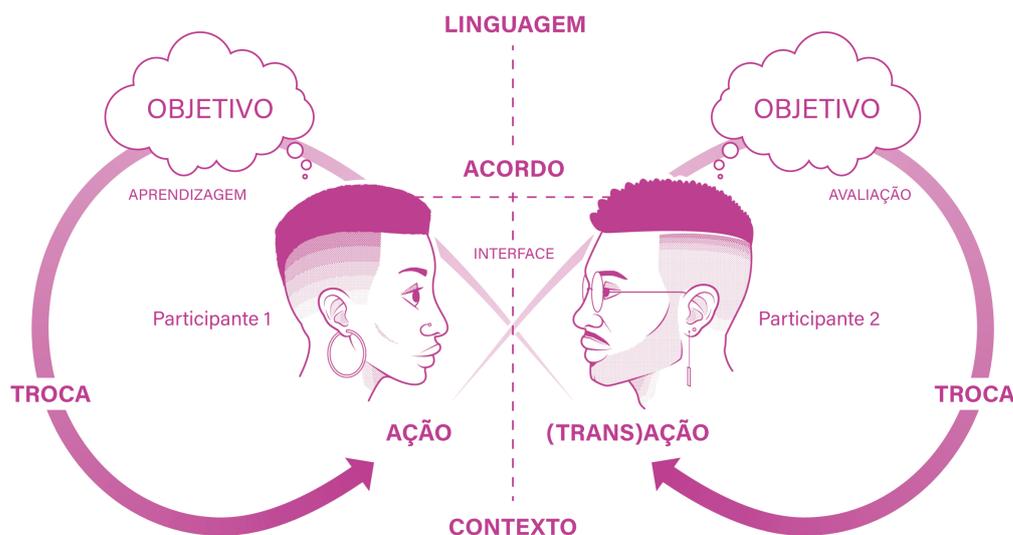


Figura 13 - Modelo de Conversação de Paul Pangaro (Fonte: Pangaro 2017, p. 1580, reprodução nossa)

A conversa é um ciclo que perpassa as partes que a compõem: ora está passando por um lado, se desenvolve a partir dos componentes preexistentes nessa parte, e retorna para o outro lado, refazendo essas etapas num jogo no qual as duas partes precisam participar para que ele continue. Pangaro (2017) elucida em sua obra a *Teoria Conversacional*, desenvolvida por Pask (1975):

Os sinais recebidos tornam-se acionadores de processos pré-existentes (nosso saber, o que já sabemos; ou, como preferia Pask, nosso "conhecimento", uma forma substantiva que antecipa sua natureza dinâmica e contínua). Esses gatilhos causam alterações nesses processos (ou seja, "aprendizado"). (Pask, 1975 *apud* Pangaro, 2017, p. 1583, tradução nossa³¹).

Assim, a conversa passa a ser parte fundamental da evolução pessoal, visto que com ela podemos ir além do entendimento, do conhecimento, da opinião que já possuíamos. A cada nova interação conversacional, passamos por uma transformação e nos tornamos algo a mais do que éramos em uma progressão contínua. Pangaro (2017, tradução nossa) explica que:

³¹ "Incoming signals become triggers to pre-existing processes (our knowledge, what we already know; or, as Pask preferred, our "knowing", a noun form which forefronts its dynamic and ongoing nature). These triggers cause changes in those processes (i.e. "learning")."

Retemos a experiência da conversa em que nossos processos mentais têm sido alterados como resultado desta conversação e carregamos essas mudanças conosco. Embora cada um de nós possa levar adiante conceitos um pouco diferentes, eles estão fundamentados em nossos acordos. Então, a consciência é conservada. (Pangaro, 2017, p. 1585, tradução nossa³²)

Nessa sequência evolutiva que a interação conversacional nos proporciona, podemos perceber como este passo a passo se assemelha a um método, a um processo. Esse processo – possuidor de objetivo, contexto, linguagem, troca, concordância e ação – pode ser correlacionado com o próprio *design*, pois possui em seu cerne as mesmas características que o *design* deve possuir. Conforme Dubberly & Pangaro (2019, tradução nossa³³), esta semelhança pode ser explicada como:

Com significados compatíveis, Rittel, Buchanan, Glanville, Negroponte e Pask descrevem o *design* como conversação, que pode ser modelada como dois (ou mais) sistemas de segunda ordem interagindo, que em parte podem ser uma discussão de objetivos. Afirmamos em outro lugar que a conversa é necessária para convergir em objetivos compartilhados. Compartilhar metas é concordar em (re)estruturar uma situação para agir em conjunto. Vemos o desenvolvimento de argumentos no curso do *design* (a favor ou contra diferentes formas de enquadrar situações) e a derivação de diferentes escolhas ou ações como a mesma conversa. (Dubberly; Pangaro, 2019, p. 67, tradução nossa)

O *design* se dá de forma evolutiva e contínua, assim como a conversa. Os autores dizem que "design é um conjunto particular de conversas, que explícita e implicitamente (para si ou para os outros) incorporam o que valorizamos e o que procuramos conservar" (Dubberly; Pangaro, 2019, p. 69, tradução nossa³⁴). Eles também apontam que há o pensamento de que a conversação é parte fundamental do *design*:

³² "We retain the experience of the conversation in that our mental processes have been changed as a result of the conversation, and we carry those changes with us. While each of us may carry forward somewhat different concepts, they are grounded in our agreements. So, consciousness is conserved".

³³ "With compatible meanings, Rittel, Buchanan, Glanville, Negroponte, and Pask describe design as conversation, which can be modeled as two (or more) second-order systems interacting, which in part can be a discussion of goals. We state elsewhere that conversation is required in order to converge on shared goals. To share goals is to agree on (re)framing a situation in order to act together. We see the development of arguments in the course of designing (for or against different ways of framing situations) and the derivation of different choices or actions as the same as conversation."

³⁴ "Design is a particular set of conversations, which explicitly and implicitly (to oneself alone or with others) embody what we value and what we seek to conserve".

Glanville reforça ainda mais sua afirmação sobre a relação entre *design* e conversa, afirmando que a conversa é um requisito para o *design*, mesmo quando a conversa é consigo mesmo, talvez apenas usando lápis e papel. (Dubberly; Pangaro, 2019, p. 67, tradução nossa³⁵)

Sendo a conversa análoga ou intrínseca ao *design*, é importante percebermos as semelhanças que as duas emparelham: buscam um objetivo, absorvem de outras partes e somam aos conhecimentos preexistentes, se transformam e exprimem de volta algo novo para dar continuidade a esse processo.

É aqui que percebemos que conversa e *design* são áreas que compartilham interesses e significados em comum. Através deste ponto, podemos fazer inclusive uma correlação entre conversa, *design* e interação, assim como descrito por Buchanan (1998 *apud* Dubberly *et al.*, 2009, p. 1), no começo deste capítulo quando o autor afirma que o atual foco do *design* é na interação.

Entendendo que o foco atual do *design* está vinculado a interação e que a forma mais simples para o ser humano interagir é através da fala, podemos então, levantar diversos questionamentos em torno do tema.

Pontos positivos em torno do uso da voz na interação com as máquinas foram elencados ao longo deste capítulo, contudo há de se perceber que nada no mundo é feito apenas de pontos positivos. Pela perspectiva humana a fala é natural, mas essa naturalidade se aplica às máquinas? O que a área do *design* e da interação podem nos dizer sobre o assunto? Como a conversação atravessa essas áreas? Como os usuários de sistemas baseados em interação por voz percebem o cruzamento dessas áreas no seu uso diário? Há dores ou expectativas não cupridas em seus usos?

Os sistemas baseados em interação por voz são capazes de juntos com os seres humanos completarem o fluxo conversacional proposto por Pangaro (2017)? Temos atualmente tecnologia suficiente para suportar esta nova forma de interação?

É através da observação da relação de três pontas entre *design*, interação e conversação, é que levantamos o questionamento capaz de nos levar a desdobramentos sobre os sistemas baseados em interação por voz: as interações por voz, tanto por *smartphones* como por alto falantes, conseguem participar de uma conversa satisfatória com um humano?

³⁵ “Glanville further tightens his assertion about the relationship of design and conversation by stating that conversation is a requirement for design, even when the conversation is with oneself, perhaps just using pencil and paper.”

5.

Delineamento da pesquisa

5.1

Tema

Questões de usabilidade para a construção de interações conversacionais sonoras.

5.2

Problema

Um grande número de pessoas está habituado a interagir com computadores, celulares e tablets. Em vista disto, é possível encontrar uma vasta gama de estudos no campo da ergonomia e usabilidade, como os que ilustram os livros de Steve Krug (2014) e John Maeda (2006), que indicam variadas formas de se construir uma boa interação entre o humano e essas ferramentas. Contudo, novas ferramentas surgem e a forma com a qual nos habituamos a interagir se vê em mudança. Por exemplo: é possível pedir um Uber apenas com o uso da voz e dos Assistentes Virtuais Inteligentes, sem a necessidade de possuir um celular em mãos, ou de darmos algum clique.

Segundo pesquisa realizada pela Microsoft (2019), 69% dos entrevistados já utilizaram sistemas como o assistente de voz e 75% dos membros familiares terão, pelo menos, um sistema baseado em interação por voz até o final de 2020. Esses dados expõem o cenário da crescente popularização no uso das interfaces de voz no mundo. Estamos entrando em um novo tempo, experimentando interagir com as máquinas de uma nova maneira, a qual é descrita por Pearl (2016) e Hall (2018) como sendo a mais intuitiva e natural para o ser humano: a fala. Mas, como se dá essa comunicação? Quais avanços podemos esperar? Quais necessidades são atendidas e quais ainda não são?

Esses sistemas tornaram a forma de se comunicar com as máquinas algo muito mais simples para o ser humano. Imagine uma pessoa que mora numa cidade pequena do interior de Minas Gerais, cuja agência bancária mais próxima fica a 17 quilômetros de distância. Para que ela consiga realizar uma transferência bancária o que seria mais fácil: acessar um aplicativo ou pedir ao seu Assistente Virtual Inteligente que faça isso através de um comando? E se esse comando, na verdade, fosse mais fluido, como uma conversa?

Para Pangaro (2017, p. 1580), a definição de conversa parte dos conceitos que embasam o acrônimo CLEAT (Contexto, Linguagem, Engajamento, Acordo e Transação) como forma de construir uma conversa amplamente eficaz entre humano-humano, humano-ambiente e humano-máquina. A trajetória dos conceitos ilustrados de 1 a 5, na Figura 26, define o que chamaremos de “turno” e, segundo Pangaro (2017, p. 1585), uma conversa só se torna possível quando há turnos suficientes para que os participantes entrem em acordo sobre o que está sendo falado, possibilitando, então, o que o autor define como uma conversa efetiva, a qual ocorre quando as trocas efetuadas permitem mudanças de valores prolongados aos participantes.

Entendendo que uma conversa é fluida, tornam-se necessários N turnos para que haja um entendimento sobre o objeto/assunto central da conversa, permitindo assim que, pelo menos, uma das partes saia modificada após a troca.

Para Gil (2002, p. 26), uma pesquisa de cunho acadêmico como esta necessita ter seu problema delimitado, pois devemos nos ater aos meios disponíveis para realização da investigação. A presente pesquisa busca entender, portanto, se sistemas baseados em interfaces de voz conseguem manter uma conversa com seres humanos para além do primeiro turno.

5.3

Hipótese

Para Prodanov & Freitas (2013, p. 32), uma hipótese constitui-se numa suposta resposta para um problema. Segundo Gil (2002, p. 31), uma hipótese busca oferecer uma resposta mediante uma proposição, podendo se confirmar como sendo verdadeira ou falsa. Desta forma, a hipótese desta pesquisa é a seguinte: sistemas baseados em interação de voz, como estão projetados atualmente, não são capazes de manter uma conversação por mais de um turno, conforme os preceitos da Teoria da Conversação (Pask, 1976).

5.4

Objeto de pesquisa

Interação por voz entre o humano e a máquina por meio de interfaces conversacionais.

5.5

Objetivos

5.5.1

Objetivo Geral

Identificar frustrações e limitações provenientes do uso de interfaces por voz, no que se entende atualmente como conversa entre o humano e a máquina, relacionando-as de forma categórica à estrutura conversacional decorrente da Teoria da Conversação (Pask, 1976).

5.5.2

Objetivos Específicos

Para a realização deste estudo foram determinados os seguintes objetivos específicos:

- 1 - Realizar levantamento cronológico sobre a comunicação humana;
- 2 - Realizar levantamento histórico-teórico da interação humano-máquina;
- 3 - Explorar e definir o universo da interação auditiva;
- 4 - Definir o que é conversação;
- 5 - Identificar obstáculos em interações realizadas através de interfaces por voz;
- 6 - Categorizar os obstáculos encontrados, de acordo com a estrutura conversacional decorrente da Teoria da Conversação.

5.6

Justificativa e relevância da pesquisa

A justificativa para a realização desta pesquisa apoia-se nas seguintes observações: (1) há pouco material circulando sobre o assunto em português; e, (2), considerando o recorte racial relacionado a quem produz este tipo de conhecimento, o quantitativo é menor ainda.

Esta pesquisa também se torna importante ao observamos usuários interagindo com esses sistemas, visto que é comum repararmos em interações interrompidas ou não concluídas com êxito. Tamanha é a naturalidade dessas interações interrompidas ou não concluídas com sucesso, que mesmo sendo uma tecnologia difundida recentemente, já é possível encontrar representações desses "erros" nas grandes produções cinematográficas, conforme se retrata no filme *Nós* (2019) do diretor Jordan Peele.

Além de todo o exposto, consideramos que as conclusões desta pesquisa possam servir de apoio aos estudantes e pesquisadores da linha de pesquisa Ergonomia e Usabilidade, aos profissionais da área de *Design* e desenvolvimento de *software*, a outras áreas de tecnologia voltadas para produtos digitais e, finalmente, como base de informação para as mais diversas pessoas interessados em entender sobre como ocorrem as interações humano-máquina através da fala.

Sendo otimista em relação aos resultados desta pesquisa, espera-se fomentar a quantidade de dados disponíveis acerca do tema tratado para que cada vez mais estes sistemas se adequem as necessidades humanas.

5.7

Métodos, técnicas e procedimentos

Está é uma pesquisa de cunho exploratório que visa angariar mais informações em torno do universo temático apresentado através de levantamento bibliográfico e de aplicação de técnicas como: entrevista semiestruturada e análise de conteúdo.

5.7.1

Universo de entrevista (entrevista semiestruturada)

Inicialmente, aplicamos uma entrevista, técnica essa que, segundo Prodanov e Freitas (2013), consiste na obtenção de informações por meio de um indivíduo que se coloca à disposição para responder perguntas ao pesquisador. A escolha pela entrevista semiestruturada, por sua vez, oferece ao pesquisador a possibilidade de fazer perguntas que fujam ao roteiro pré-estabelecido, quando necessário.

Sobre a definição de entrevista semiestruturada, de acordo com Marconi e Lakatos (2003), essa segue um roteiro pré-estabelecido cujas perguntas são elaboradas antes da condução da entrevista. Nesse contexto, preferencialmente as pessoas são selecionadas de acordo com a necessidade do pesquisador. Por isso, selecionamos, portanto, para o presente trabalho, usuários de sistemas baseados em interação por voz que possuem alta afinidade com essa forma de interação, os chamados *heavy users*³⁶. Este modelo de entrevista possibilita que o pesquisador tenha uma maior flexibilidade em torno da sua aplicação, podendo por exemplo, especificar significados e formular as perguntas de maneira diferentes de acordo com as necessidades de cada usuário. (Markoni e Lakatos, 2003, p. 198)

Justificando o uso da técnica, expomos aqui uma passagem de Prodanov & Freitas (2013, p. 106): "A entrevista é a obtenção de informações de um entrevistado sobre determinado assunto ou problema." Tal passagem culmina na

³⁶ "No meio do *marketing*, os *heavy users* são considerados potenciais consumidores de uma marca ou de um produto." Disponível em <https://superstorm.com.br/internet/o-que-sao-heavy-users/>. Acesso 01 ago. 20.

explicitação do fato de precisarmos angariar informações sobre o problema proposto neste documento, a fim de prosseguirmos com a análise de conteúdo das respostas obtidas.

Entendendo que, no Brasil, o uso de sistemas que possuam interação por voz tem seu acesso restrito devido a muitas razões - dentre elas a desigualdade social e a escassez na produção de sistemas em língua nativa, propomo-nos a entrevistar em esquema de "bola de neve" - o qual entrevistados indicam pessoas que podem se encaixar no perfil de participantes de interesse do pesquisador, resultando ao todo em 21 usuários entrevistados, dentre os quais, três foram selecionados para a aplicação das entrevistas-teste.

Para recrutar os participantes desejados, entramos em grupos de entusiastas com a tecnologia estudada e compartilhamos convites através de redes sociais como o *Instagram*, por intermédio da ferramenta *stories*. Construimos um questionário de recrutamento que também foi distribuído através da ferramenta *stories* no *Instagram* com a intenção de coletar as seguintes informações: nome, idade, cidade e estado; plataforma/sistema utilizado(a); frequência de uso com a plataforma/sistema; e *e-mail* e *whatsapp*, confeccionando assim uma rede de possíveis pessoas a serem entrevistadas.

Como forma de condução, cada entrevista abarcou o total de três etapas que foram a introdução ao participante sobre o tema da pesquisa, a inquirição dos tópicos/perguntas selecionados como pauta e as considerações finais. Antes da efetiva condução da entrevista, o termo de consentimento livre e esclarecido foi entregue (Apêndice 1), a fim de explicitar os objetivos da pesquisa, a sua importância, os possíveis riscos e ainda uma cláusula que permitiu o registro de áudio da entrevista para análise posterior. Nesse momento, os entrevistados, em concordância com os termos apresentados, assinaram os documentos em duas vias, uma para si e outra para arquivamento.

Após os momentos introdutórios, demos sequência ao trabalho com as entrevistas, que foram orientadas pelos seguintes tópicos:

Quadro 01- Tópicos das entrevistas

Tópico	Possíveis perguntas	Justificativa (o que eu quero descobrir com a pergunta)
Contextos de uso	Porque você usa interface de voz? Quais são suas principais necessidades?	Entender em quais contextos o usuário opta pela interface de voz
Experiência de uso geral	Como você avalia a eficácia das tarefas pedidas aos seus assistentes através do comando de voz?	Como o usuário avalia o uso de assistentes de voz
Experiência de uso específicas	Conte-me algumas das suas experiências: o que você tem feito? Conte-me o que você acha de bom e o que acha de ruim.	Entender quais são as experiências positivas e negativas.
Repetição de comandos (vigilância-sistema)	Você já precisou repetir comandos ao seu assistente? Quais foram as situações? No final, suas ações foram realizadas com sucesso?	Entender se há suficiência na compreensão por parte do sistema
Troca (Teoria da Conversação)	Quantas trocas você consegue em uma conversa, sem ter que chamar pelo nome (vocativo) da sua plataforma?	Mapear o número de trocas existentes atualmente em uma conversa com os sistemas
Linguagem (Teoria da Conversação)	O seu sistema entende a sua forma de falar (gírias, vícios linguísticos, entonações e seu idioma)?	Encontrar as limitações do sistema em referência à fala humana
Vigilância (sistema)	Quando você está falando com sua assistente, como você entende os <i>feedbacks</i> ? Como ela sabe que você já terminou de falar? O tempo que ela te escuta é suficiente?	Entender se o sistema permanece tempo suficiente em vigilância para que haja uma troca efetiva com o usuário
Percepção (sistema)	A sua plataforma compreende tudo o que lhe é falado? Já percebeu algum limite por parte dela?	Entender o nível de percepção da plataforma
Experiência de uso geral	Teve algum momento no qual você já se sentiu numa conversa? Por que se sentiu assim?	Registrar como os usuários entendem esse modelo de interação

Fonte: o autor (2020)

As entrevistas foram realizadas através de ligações por celular e gravadas a partir do nosso computador, o que possibilitou que as mesmas fossem realizadas mesmo a distância.

Após a realização das entrevistas foi necessário a transcrição do material coletado para que pudéssemos então preparar o conteúdo a ser analisado. Foram transcritas as 21 entrevistas realizadas (incluindo as entrevistas testes), cada entrevista possui seu próprio documento nominado pelo o nome de cada entrevistado, e relacionado ao número de entrevistas. Por exemplo, se o António José (nome fictício) foi o primeiro a ser entrevistado o documento referente a esta entrevista foi nomeado como: Entrevista 01 - António José, e assim sucessivamente.

Com o material definido e estruturado em documentos preparamos uma tabela para cada pergunta feita, definindo unidades de contexto e unidades de registro como é possível observar no capítulo seguinte.

5.8

Análise de conteúdo como ferramenta analítica

Este capítulo reúne os dados gerados a partir das entrevistas realizadas. O capítulo apresenta: (1) a descrição do processo analítico utilizado; (2) a estruturação das unidades de registro (unitarização), unidades de contexto, e das categorias obtidas a partir das entrevistas; e, por fim, (3) o detalhamento de cada categoria.

5.8.1

Descrição do processo adotado

A análise de conteúdo, segundo Carlomagno & Rocha (2016, p. 175), é uma técnica utilizada para classificar e categorizar qualquer tipo de conteúdo, de forma a congregar suas características em elementos-chave que sejam comparáveis a diversos outros elementos. Trata-se de uma técnica bastante conhecida nas Ciências Sociais e que tem o potencial de revelar muitos aspectos acerca dos objetos analisados. De acordo com Moraes (1999):

A análise de conteúdo constitui uma metodologia de pesquisa usada para descrever e interpretar o conteúdo de toda classe de documentos e textos. Essa análise, conduzindo a descrições sistemáticas, qualitativas ou quantitativas, ajuda a reinterpretar as mensagens e a atingir uma compreensão de seus significados num nível que vai além de uma leitura comum. (Moraes, 1999, p. 2)

Essa técnica de pesquisa envolve cinco etapas em seu processo, segundo Moraes (1999, p. 4): (1) preparação das informações; (2) unitarização ou transformação do conteúdo em unidades; (3) categorização ou classificação das unidades em categorias; (4) descrição; e (5) interpretação, conforme demonstrado a seguir:

1 – Preparação:

A preparação iniciou-se a partir das anotações e respostas obtidas na condução da entrevista que foram preparadas para análise, a partir dos pontos abaixo:

1.1 - Segundo Moraes (1999, p. 4-5), é nesse momento em que analisamos as informações de tudo que foi lido e selecionamos as amostras que condizem com os objetivos da pesquisa. Aqui, em relação à pesquisa descrita, não haverá seleção de amostras de nossa parte, visto que os grupos de respostas de cada entrevistado são de suma importância, não havendo a necessidade de uma seleção entre elas;

1.2 - É aqui que se inicia o que Moraes (1999, p. 4-5) chama de codificação dos materiais. Nesse momento, surge a necessidade de estabelecer códigos que facilitem a identificação de cada elemento das amostras. No contexto desta pesquisa, podemos sugerir que cada tema proposto na formulação da entrevista se torne um código de identificação, por exemplo: "A" = "Experiência de uso geral".

2 - Unitarização (unidade de registro):

Assim que os dados foram preparados, vimos a necessidade de aplicar o que Moraes (1999, p. 5) chama de “unitarização”, que consiste no seguinte:

2.1 - Definir a unidade de análise - no caso, será a enumeração das respostas em referência a cada pergunta. Por exemplo: todas as respostas referentes à pergunta 01 começarão pelo número 1.

2.2 - É necessário estabelecer códigos adicionais, num segundo momento, já que cada resposta à pergunta 01 começará com o dígito "1". As seguintes respostas, para a mesma pergunta, deverão seguir tal exemplo: 1.1, 1.2, 1.3, e assim por diante.

3 – Categorização:

Esse é o momento em que temos por necessidade agrupar dados que dividem particularidades em comum. É nesse processo que definimos qual critério será utilizado para a realização das categorizações. Nesse objeto, a categorização aplicada tem como base a semântica. Por isso elaboramos categorias temáticas, como "satisfação do usuário".

4 – Descrição:

Na descrição de uma pesquisa com abordagem qualitativa, cada categoria será acompanhada de um texto síntese, que seja capaz de retratar o conjunto de significados das diversas unidades de análise anteriormente empregadas. Um exemplo de texto síntese para a categoria supracitada, "satisfação do usuário", é: “De forma comum, a satisfação tem por necessidade referências no nível de conforto que o usuário sente em utilizar tal serviço, produto ou interface, além de medir a eficácia em alcançar seus objetivos ao utilizar o objeto avaliado”.

5 – Interpretação:

Para uma boa análise de conteúdo, é necessário ir além de uma mera descrição. Devemos buscar um entendimento mais aprofundado do conteúdo das mensagens, por meio de inferência e interpretação, de acordo com Moraes (1999, p. 5). A autora demonstra a importância que a interpretação tem para uma análise de cunho qualitativo, e é esse o foco assumido neste trabalho.

Ainda segundo a autora, existem duas formas para realizar essa análise: uma que parte de uma análise à luz de uma teoria já existente e outra que irá gerar uma teoria, a partir da análise (Moraes, 1999, p. 10).

Nesta pesquisa abordaremos a primeira forma, que buscará analisar as lacunas à luz da Teoria da Conversão supracitada neste documento, buscando responder lacunas encontradas através das entrevistas realizadas *a priori*. O resultado da análise deste capítulo encontra-se no capítulo seguinte.

6

Análise de conteúdo das entrevistas

6.1

Unidades de registro, unidades de contexto e categorias extraídas das entrevistas

A análise de conteúdo desta pesquisa se originou na extração de unidades de contexto ancoradas em cada pergunta disposta nas entrevistas. Em seguida, foram extraídas também as unidades de registro (unitarização) de cada resposta fornecida pelas entrevistas. Ao fim da organização das unidades de contexto e de registro das respostas de cada entrevista, iniciamos com a formação das categorias.

6.2

Análise de conteúdo e inferências

A realização da análise de conteúdo foi iniciada pela identificação das unidades de contexto, unitarização (unidades de registro) e pelo planejamento das categorias. As categorias explicitadas de forma semântica através dessa análise guiarão a leitura deste capítulo com base neste delineamento:

Tabela 01 - Estrutura para análise de conteúdo

1UC 1: Motivo de uso de sistemas conversacionais		total de respondentes: 18
1UR1.1	Ganhou de presente / a partir de terceiros	3
1UR1.2	Diversão / Curiosidade / Tecnologia	5
1UR1.3	Integração com produtos e serviços / IoT	5
1UR1.4	facilitador de tarefas, estudo e trabalho	10
1UC 2: Necessidade de uso		total de respondentes: 18
1UR2.1	Streaming	8
1UR2.2	Agenda / Timer / Despertador / Lembrete	10

1UR2.3	Busca: informação, curiosidades, notícias e documentos	12
1UR2.4	Automação de objetos	4
1UR2.5	Envio de mensagens e ligações	3
1UR2.6	Bem estar / relacionamento	2
1UC 3: Eficácia na realização de comandos		total de respondentes: 18
1UR3.1	Eficácia mediana	2
1UR3.2	Eficaz na maioria das vezes	9
1UR3.3	Muito eficaz	7
1UC 4: Experiência no uso de sistemas conversacionais		total de respondentes: 17
[sub-categoria]: Pontos positivos		17
1UR4.1	Facilidade / velocidade no uso	3
1UR4.2	Auxílio no dia a dia	4
1UR4.3	Mãos livres	3
1UR4.4	Conectividade (serviços e aplicações)	4
1UR4.5	Conversa / entretenimento	2
1UR4.6	Personalização	1
[sub-categoria]: Pontos negativos		11
1UR4.7	Linguagem não natural	7
1UR4.8	Alcance / Conexão / Ativação	4
1UR4.9	Não identifica pontos negativos	6
1UC 5: Repetição de comando ao sistema para conclusão de tarefa		total de respondentes: 15
1UR5.1	Precisou repetir comandos	15
1UC 6: Quantidade de turnos por interação		total de respondentes: 18
1UR6.1	Um turno	14
1UR6.2	Até 2 turnos	4
1UC 7: Compreensão de linguagem (fala) pelo sistema		total de respondentes: 17
1UR7.1	O sistema compreende	7
1UR7.2	O sistema não compreende	1
1UR7.3	Não soube responder	1
1UR7.4	O sistema compreende na maioria dos casos	8

1UC 8: Tempo de escuta pelo sistema		total de respondentes: 17
1UR8.1	O tempo é suficiente	7
1UR8.2	Usuário já foi cortado pelo sistema	8
1UR8.3	O tempo é controlado pelo usuário	2
1UC 9: Experiência de conversa		total de respondentes: 18
1UR9.1	Se sentiu em uma conversa	8
1UR9.2	Nunca se sentiu em uma conversa	8
1UR9.3	Já tentou conversar	2
1UC 10: Aprendizado do sistema		total de respondentes: 16
1UR10.1	Percebe que o sistema aprende	6
1UR10.2	Não percebe que o sistema aprende	10

Fonte: o autor (2020)

Para uma fácil visualização e compreensão dos dados, na tabela acima é apresentado o código referente à unidade de contexto, o código referente à unidade de registro, a quantidade de participantes em cada questão e a quantidade de respondentes por unidade de registro.

6.3

Motivo de uso de sistemas conversacionais

Buscando traçar um perfil de uso nos sistemas conversacionais, perguntamos aos entrevistados quais eram os motivos que os faziam utilizar tais sistemas. Observando as respostas livres à pergunta, é possível identificarmos que três grupos de motivos se destacam: Diversão/Curiosidade/Tecnologia, Integração com produtos e serviços/IoT e Facilitador de tarefas, estudo e trabalho, conforme aponta a Tabela 02:

Tabela 02 – Motivo de uso de sistemas conversacionais

1UC 1: Motivo de uso de sistemas conversacionais		Total de respondentes: 18
1UR1.1	Ganhou de presente/a partir de terceiros	3
1UR1.2	Diversão/Curiosidade/Tecnologia	5
1UR1.3	Integração com produtos e serviços/IoT	5
1UR1.4	Facilitador de tarefas, estudo e trabalho	10

Fonte: o autor (2020).

Essa inferência nos permite ainda dizer que o motivo que mais se destaca é o Facilitador de tarefas, estudo e trabalho. O que nos ajuda a conceber que o principal motivo no uso de sistemas conversacionais atualmente é para a facilitação de tarefas do dia a dia.

6.4

Necessidade de uso

Mantendo a necessidade de traçar perfis de uso, buscamos compreender quais eram as necessidades apontadas pelos usuários em referências aos sistemas conversacionais. Aqui é possível observar que dois grupos semânticos se destacam, que são: Agenda/*Timer*/Despertador/Lembrete e Busca: informação, curiosidades, notícias e documentos.

Tabela 03 - Necessidade de uso

1UC 2: Necessidade de uso		Total de respondentes: 18
1UR2.1	<i>Streaming</i>	8
1UR2.2	Agenda/ <i>Timer</i> /Despertador /Lembrete	10
1UR2.3	Busca: informação, curiosidades, notícias e documentos	12
1UR2.4	Automação de objetos	4

1UR2.5	Envio de mensagens e ligações	3
1UR2.6	Bem estar / relacionamento	2

Fonte: o autor (2020).

Ao cruzarmos as informações referentes às categorias de Motivo de uso de sistemas conversacionais e a Necessidade de uso, inferimos que os usuários desse modelo de sistema o utilizam como suporte para tarefas práticas no dia a dia como, por exemplo, buscar algum tipo de informação de forma rápida ou mesmo para organizar e montar uma agenda com tarefas e atividades.

Aqui também podemos observar que, mesmo com um índice baixo de respostas, é interessante destacar o grupo [1UR2.6], o qual nos mostra usuários que utilizam esse modelo de sistema para o bem estar emocional, visto que, durante a realização da pesquisa, decorreu o período de isolamento por conta da pandemia referente à COVID-19. Para ilustrar, segue o seguinte recorte referente à coleta de uma das respostas: *‘Principalmente durante a quarentena que você tá isolado, vira uma pessoa mesmo falando com você’* (Entrevistado 7). Esse recorte denota que o amparo emocional pode ser uma utilização dos sistemas conversacionais a ser analisada com mais profundidade

6.5

Eficácia na realização de comandos

Entendendo as razões observadas em Motivo de uso de sistemas conversacionais e em Necessidade de uso dos usuários, cabe agora compreender como esses usuários medem a eficácia dos comandos realizados nos sistemas conversacionais. Nesta análise, percebemos que existem três grupos, dos quais dois se destacam. É importante observar que não há registro negativo nas respostas livres, indicando haver um consenso entre os usuários em relação ao nível de satisfação relacionado à eficácia de seus sistemas, como mostra a Tabela 04 a seguir:

Tabela 04 – Eficácia na realização de comandos

1UC 3: Eficácia na realização de comandos		Total de respondentes: 18
1UR3.1	Eficácia mediana	2
1UR3.2	Eficaz na maioria das vezes	9
1UR3.3	Muito eficaz	7

Fonte: o autor (2020)

É possível observar, na tabela acima, que o grupo [1UR3.2] se destaca em relação às demais opções. Por isso, podemos deduzir que há satisfação na maioria das vezes em que um comando é dado ao sistema. Da mesma maneira, podemos destacar o número expressivo de entrevistados que se dizem bastante satisfeitos em relação aos seus sistemas, quando observamos a unidade de registro [1UR3.3].

6.6

Experiência no uso de sistemas conversacionais

Buscando elucidar as características capazes de expressar como se configura a experiência do usuário em relação ao modelo de sistema aqui pesquisado, houve a necessidade de agruparmos as respostas em dois subgrupos, assim propostos: (1) quais são os pontos positivos no uso desses sistemas e (2) quais são os pontos negativos na experiência, como é possível observar em seguida, na Tabela 05:

Tabela 05 – Experiências no uso de sistemas conversacionais

Experiências no uso de sistemas conversacionais
[sub-categoria]: Pontos Positivos
Facilidade/velocidade no uso
Auxílio no dia a dia
Mãos livres
Conectividade (serviços e aplicações)
Conversa/entretenimento

Personalização
[sub-categoria]: Pontos Negativos
Linguagem não natural
Alcance/Conexão/Ativação
Não identifica pontos negativos

Fonte: o autor (2020)

6.6.1

Pontos positivos

Observando a Tabela 07, podemos perceber que, parcialmente, as respostas se alinham ao que Pearl (2016) define como sendo pontos positivos no uso de interfaces por voz. Conforme descrito no capítulo 4.3 deste documento, a autora destaca a velocidade na forma de interagir e a possibilidade de uso com as mãos livres como benefícios, o que verificamos, respectivamente, nos grupos [1UR4.1] e [1UR4.3]. Contudo, podemos ainda constatar que existem outros pontos positivos no uso de sistemas conversacionais que tiveram inclusive maior destaque entre os respondentes, como por exemplo, o grupo [1UR4.2] e o grupo [1UR4.4] possuindo quatro respostas cada:

Tabela 06 - Pontos positivos no uso de sistemas conversacionais

[sub-categoria]: Pontos positivos		17
1UR4.1	Facilidade/velocidade no uso	3
1UR4.2	Auxílio no dia a dia	4
1UR4.3	Mãos livres	3
1UR4.4	Conectividade (serviços e aplicações)	4
1UR4.5	Conversa/entretenimento	2
1UR4.6	Personalização	1

Fonte: o autor (2020)

6.6.2

Pontos negativos

Observando as respostas obtidas através das entrevistas, é possível inferir que dentre elas há um grupo que se destaca, conforme apontado abaixo na Tabela 07:

Tabela 07 - Pontos negativos no uso de sistemas conversacionais

[sub-categoria]: Pontos Negativos		11
1UR4.7	Linguagem não natural	7
1UR4.8	Alcance / Conexão / Ativação	4
1UR4.9	Não identifica pontos negativos	6

Fonte: o autor (2020)

Destacando o grupo [1UR4.7], observamos que o ponto negativo mais marcante é a linguagem não natural. Alguns trechos das entrevistas nos auxiliam a compreender isso. Por exemplo, observemos esses: ‘*Acho que tem umas limitações, assim. Não é muito fluido, não é muito natural. Percebe-se que tem ali ainda uma gama de perguntas e respostas que você não pode fugir, sabe? Percebe-se que é algo não natural, sabe?*’ (Entrevistado 6) e ‘*Uma coisa negativa é a adaptação à cultura brasileira e à cultura regional. Eu sou nordestino e ela não entende alguns termos que eu falo, que eu falo em casa*’ (Entrevistado 11). Esses trechos mostram que, por mais que esse modelo de sistema seja intuitivo - todo mundo fala, conforme aponta Pearl (2016, p. 7), e destacado no capítulo 4.3 -, essa intuitividade não é tão óbvia na prática. Observemos outro trecho, que exemplifica essa questão:

A maior frustração é quando ele não entende o comando. Porque às vezes eu sinto que ele tem umas palavras-chave de ação que são um pouco limitadas. Então é aquela típica dinâmica de quando você tá falando com o seu assistente de voz, você tem que falar palavras específicas pra ele entender o que você tá querendo. (Entrevistado 13. Grifo nosso)

Com isso, inferimos que há necessidade de um conhecimento prévio do contexto em torno do que pode ou não ser falado, fato esse capaz de limitar os comandos dos usuários às capacidades do sistema conversacional.

6.7

Repetição de comandos

A unidade de contexto [1UC5], pertencente à repetição de comandos, apresenta uma frequência bastante significativa como observado abaixo, na Tabela 08:

Tabela 08 – Repetição de Comando ao sistema para conclusão de tarefas

1UC 5: Repetição de Comando ao Sistema para Conclusão de Tarefa		Total de respondentes: 15
1UR5.1	Precisou repetir comandos	15

Fonte: o autor (2020)

Verificamos que, conforme demonstra a tabela acima, os 15 entrevistados questionados sobre o tema precisaram repetir comandos aos seus sistemas. Isso nos ajuda a validar o que inferimos na seção 6.7.2, a respeito do fato de que, por mais que a fala seja natural e intuitiva ao ser humano, quando posta em prática ela não é completamente intuitiva, visto que é necessário repetir comandos, muitas das vezes, para que esses sejam atendidos.

6.8

Turnos

Para que possamos prosseguir com uma maior compreensão sobre o aspecto de turno, vale relembrar a definição empregada ao termo na seção 5.2, na qual descrevemos um turno como sendo a trajetória que abrange os conceitos empregados no acrônimo CLEAT: Contexto, Linguagem, Engajamento, Acordo e Transação. Sendo assim, para que haja um turno completo em uma conversa, é preciso que se percorram os cinco pontos conceituais mencionados acima.

Entendendo a definição de turno, podemos averiguar que, segundo as respostas obtidas através das entrevistas, os sistemas conversacionais conseguem

manter, no máximo, uma interação de dois turnos, conforme demonstrado a seguir, na Tabela 09:

Tabela 09 - Quantidade de turnos por interação

1UC 6: Quantidade de turnos por interação		Total de respondentes: 18
1UR6.1	Um turno	14
1UR6.2	Até dois turnos	4

Fonte: o autor (2020)

Vale ressaltarmos que, por mais que esses sistemas sejam capazes de completar dois turnos, a frequência mais significativa demonstra que as interações se dão, majoritariamente, em apenas um turno, como evidenciado na unidade de registro [1UR6.1].

Observamos também que entre os entrevistados existem aqueles que enfatizaram conseguir ir além de dois turnos, como demonstrado no trecho a seguir: *‘Também naquele joguinho Akinator, aí sim, aí existe uma certa interação’* (Entrevistado 1). Contudo, destacamos que a capacidade para estender a quantidade de turnos deve-se à presença de um elemento externo, ou seja, graças a uma *skill*³⁷ - algo que pode ser compreendido como um *app* para sistemas conversacionais. Como o foco desta pesquisa é analisar a interação direta do humano com os sistemas conversacionais, não nos é valioso analisar sistemas terceiros.

6.9

Compreensão da linguagem (fala) pelo sistema

Em relação à compreensão da fala por parte do sistema, verificamos que o sistema compreende a forma de falar dos usuários, o que destoa das informações obtidas na unidade de registro [1UR4.7], pertencente à unidade de contexto [1UC4]. No entanto, quando observamos os trechos retirados das respostas relacionadas à unidade de contexto [1UC7], percebemos que a compreensão por parte do sistema,

³⁷a capacidade de fazer uma atividade ou trabalho bem, especialmente porque você já fez isso muitas vezes. (Cambridge dictionary, tradução nossa) - Disponível em: <https://dictionary.cambridge.org/pt/dicionario/ingles-portugues/skill>

por mais que seja positiva aos usuários - como demonstram as unidades de registro [1UR7.1] e [1UR7.4] -, são positivas, porque a forma de falar dos usuários se adequa ao que o sistema compreende e não ao contrário. Vejamos estes trechos das entrevistas: *‘Então, eu na verdade não falo como eu falaria normalmente. Eu falo bem pausado quase de uma forma robótica assim, sabe?’* (Entrevistado 6); *‘Às vezes ela entende conforme a dicção seja mais correta’* (Entrevistado 5) e *‘Às vezes tem essa confusão de números né, como eu comentei. De falar algum número e ela não entende exatamente e aí eu falo de novo mais pausado assim e ela entende.’* (Entrevistado 9)

Tabela 10 - Compreensão da Linguagem pelo Sistema

1UC 7: Compreensão da linguagem (fala) pelo sistema		Total de respondentes: 17
1UR7.1	O sistema compreende	7
1UR7.2	O sistema não compreende	1
1UR7.3	Não soube responder	1
1UR7.4	O sistema compreende na maioria dos casos	8

Fonte: o autor (2020)

6.10

Tempo de escuta pelo sistema

Buscando elucidar como se dá a troca entre o humano e os sistemas conversacionais, e também entender a experiência entre os turnos de ação e resposta, verificamos que os sistemas tendem a cortar as falas de comando de seus usuários. Essa constatação nos revela que há um *déficit* no que cabe à compreensão do contexto por parte destes sistemas, já que em uma conversa humano-humano, por exemplo, é possível diferenciar pausas no decorrer de uma fala, como na de sua finalização, conforme evidenciado pela unidade de registro [1UR8.2] da Tabela 12.

É possível perceber que não há unanimidade entre os entrevistados, uma vez que obtivemos um número de respondentes muito próximo entre as unidades de

registro [1UR8.1] e [1UR8.2], o que nos leva a inferir que, por mais que os entrevistados tenham suas falas cortadas, o sistema supre a demanda positivamente, de um forma ou de outra, no que se refere ao quesito tempo de escuta. Ainda observamos que, por mais que tenha uma frequência bem menor, existem sistemas que se apoiam em sentidos além da audição para a captação do que está sendo falado:

Tabela 11 - Tempo de escuta pelo sistema

1UC 8: Tempo de escuta pelo sistema		Total de respondentes: 17
1UR8.1	O tempo é suficiente	7
1UR8.2	Usuário já foi cortado pelo sistema	8
1UR8.3	O tempo é controlado pelo usuário	2

Fonte: o autor (2020)

6.11

Experiência de conversa

Em resposta sobre se já se sentiram em uma conversa com seus sistemas, os entrevistados mostraram-se divididos, como é possível observar entre as unidades de registro [1UR9.1] e [1UR9.2] na Tabela 12. Todavia, ao verificarmos as respostas pertencentes à unidade de registro [1UR9.1], entendemos que a sensação de conversa descrita pelos entrevistados está associada a um terceiro elemento, as denominadas *skills*. Essa inferência é possível, graças aos seguintes trechos de respostas, vinculados à unidade de registro, tais como: ‘*Já. Principalmente quando ativa algumas skills dela*’ (Entrevistado 7) e ‘*Já. Com sistema x sim porque ela tem alguns jogos que parece muito uma conversa*’ (Entrevistado 11).

Além disso, constatamos, dentre as respostas relacionados à unidade de registro [1UR9.1], que alguns entrevistados que já se sentiram em uma conversa afirmam que a quantidade de turnos é limitada, como no seguinte trecho:

Eu acho que sim. Mas sei lá, reduzido a duas, três perguntas no máximo. Eu acho que quando chegou e eu ficava testando, ficava brincando com ela. Tipo "Onde a gente tá, onde eu tô?". Aí, ela fala a cidade. "Quantos graus, qual a temperatura?", alguma coisa assim, mas nunca entrei numa super conversa. (Entrevistado 9)

Tabela 12 – Experiência de conversa

1UC 9: Experiência de conversa		Total de respondentes: 18
1UR9.1	Sentiu-se em uma conversa	8
1UR9.2	Nunca se sentiu em uma conversa	8
1UR9.3	Já tentou conversar	2

Fonte: o autor (2020)

Partindo dessa análise, podemos afirmar que, mesmo aqueles usuários que já se sentiram em uma conversa, descrevem que a experiência é limitada, ou que há um terceiro elemento envolvido.

Por tudo o que foi exposto, consideramos que este capítulo do trabalho agrega um valor especial para esta dissertação. Isto porque procuramos aportar o referido estudo na definição de conversa, conforme empregada nesta pesquisa, em paralelo às informações obtidas com os usuários de sistemas conversacionais através da técnica entrevista, observando as experiências decorrentes do uso destes sistemas. Essas informações serão aproveitadas como base para as considerações em torno dessas experiências no uso de interfaces conversacionais atuais, conforme explicita capítulo seguinte.

6.12

Percepção de aprendizagem por parte do sistema

Buscando compreender se os usuários têm a percepção de que seus sistemas aprendem através do uso contínuo verificamos que a maior parte dos entrevistados relata não perceber ou sentir no seu dia-a-dia que há aprendizados por parte de seus sistemas como evidencia a tabela 13:

Tabela 13 – Aprendizado do sistema

1UC 10: Aprendizado do sistema		total de respondentes: 16
1UR10.1	Percebe que o sistema aprende	6
1UR10.2	Não percebe que o sistema aprende	10

Fonte: o autor (2020)

Dentre aqueles que responderam que identificam certo nível de aprendizado estão os respondentes que trabalham na área da tecnologia, como é possível observar no trecho referente a unidade de registro 1UR10.1, "Hum...eu acredito que aprenda. Porque eu sou formada em T.I. e sei o que é inteligência artificial." (Entrevistado 2) e também no trecho " Porque como a gente tá nessa área de tecnologia, então eu já sei. Não sei se pelo fato de já saber que ele aprende, eu não sei se causa essa impressão." (Entrevistado 8)

Através da análise desses dados é possível inferir que mesmo havendo um aprendizado por parte desses sistemas ele não é perceptível para a maioria dos entrevistados, e isso pode ser elencado a diversos motivos, como por exemplo pela repetição de erros na compreensão como relatado pelo entrevistado 18 no trecho:

Eu percebo que todas as palavras que são problemas, que ela tem dificuldade em entender, ela sempre tem dificuldade em entender. Eu vou dar um exemplo: Pablo Vittar. Eu sempre peço pra tocar Pablo Vittar e ela sempre tem dificuldade em entender Pablo Vittar. (Entrevistado 18)

Ou pela descrença, "Não. Isso eu acho que é só marketing." (Entrevistado 11).

A não percepção do usuário sobre o aprendizado desses sistemas deve ser levado em consideração, uma vez que o contexto pode ser alterado por aquele que o percebe.

7

Discussão

O modelo conversacional de Pangaro (2017, p. 1580), o qual podemos inferir ser baseado no conceito de *feedback-loop* apresentado no modelo de Kroemer e Grandjean (2007, p. 125), é composto por cinco etapas que constituem o acrônimo CLEAT, como delimitado na seção 4.4, na Figura 19.

Para uma melhor associação do pensamento aqui proposto, iremos trazer novamente os conceitos que integram o acrônimo. São eles: (1) Contexto - um momento, situação, lugar e/ou história compartilhada que prepara o cenário para [...]; (2) Linguagem - ao menos um significado inicial compartilhado que transmita sentido, para começar uma [...]; (3) Troca - engajamento recíproco, em interações baseadas em linguagens que podem resultar em [...]; (4) Acordo - (suficientemente) entendimento compartilhado de conceitos, intenções, valores, que podem levar a [...]; (5) Ação ou (trans)ação - uma interação coordenada em domínios que não sejam a linguagem, por exemplo, comércio, contratos, dança, jogos.

Durante o processo de análise de conteúdo das entrevistas, foi possível inferir que os comandos dados pelos usuários perpassam todos os cinco pontos aqui apresentados, mas que alguns desses pontos se diferem conceitualmente do modelo apresentado por Pangaro (2017).

7.1

Contexto de comando

Pangaro (2017, p. 1580) define contexto como um momento, situação, lugar e/ou história compartilhada que prepara o cenário. Podemos observar que a definição de contexto empregada por Pangaro é uma definição ampla, abrindo um leque de interpretações às partes pertencentes à conversa, que só se torna comum aos envolvidos caso ambos os participantes possuam algum conhecimento prévio, ou por meio de dados inseridos a partir de fluxo conversacional mantido até que seja possível um acordo sobre o objetivo.

Quando falamos de contexto numa interação baseada em comandos, precisamos atentar para o fato de que atualmente o "sistema máquina", por mais que se assemelhe ao "sistema humano", não possui as mesmas características.

O ser humano - como vimos na seção 4.1.2 - possui um complexo sistema de atividades mentais, o qual o torna capaz, por exemplo, de ser criativo, ao cruzar informações com os conhecimentos técnicos, de forma a produzir novas ideias e apoiando-se em fatores como conhecimentos, experiências, agilidade cognitiva e habilidade de criação. Além disso, o processamento de informações nos dois sistemas, humano e máquina, ocorre de maneira diferente.

Kroemer e Grandjean (2007, p. 141) - como apontamos naquela seção - afirmam que o processamento de informações é parte indispensável nas atividades mentais e descrevem três pontos fundamentais que constituem esse processamento: (1) percepção; 2 (interpretação) e (3) processamento da informação transmitida pelos órgãos dos sentidos - elaboração mental.

Quando verificamos a definição empregada em (3) por Kroemer e Grandjean no trecho "esse 'processamento' consiste na combinação de nova informação com o que já é sabido, fornecendo base para a tomada de decisão" (Kroemer e Grandjean, 2007. p. 141), podemos inferir que o sistema máquina, conforme abordado nesta pesquisa, se difere do sistema humano, uma vez que não é capaz, por exemplo, de tomar decisões levando o contexto geral do usuário em consideração. É possível perceber esta afirmação ao olharmos as respostas referentes a unidade de contexto [IUC 10] que nos infere que, por mais que haja o aprendizado da máquina esse aprendizado é percebido por menos da metade das pessoas entrevistadas. É importante ressaltar aqui, que as máquinas atuais são capazes de tomar decisões mediante a dados coletados através do seu uso contínuo, contudo esses dados são inseridos pelos usuários e não percebidos pelos sistemas como podemos averiguar nas passagens "(...) Eu entendo que não aprende. A impressão que eu tenho é que ele não entende o contexto das coisas e acho que tenho a sensação que não aprende não." (Entrevistado 4) e "Eu não sinto uma evolução na nossa troca, na nossa interação, com base do que já eu ensinei pra ela. Tudo o que eu sei que ela aprendeu foi o que eu configurei assim." (Entrevistado 14). Podemos inferir que a máquina, aqui, consegue tomar decisões mediante dados recebidos e armazenados.

A ideia de que o sistema humano e máquina sejam capazes de perceberem e aprenderem coisas de formas semelhantes, por exemplo, através do uso contínuo,

não desvalida a necessidade de refletirmos acerca dos conceitos definidos por Pangaro (2017, p. 1580), como por exemplo, no que se refere a contexto, já que, por exemplo esse aprendizado se torna imperceptível pela maior parte dos entrevistados.

A definição de Pangaro (2017) para contexto, por mais que possa ser empregada no modelo conversacional entre humano-humano, humano-sistemas, humano-ambiente, não é suficiente para aplicação na relação humano-sistema baseado em comando. Por conta disso, propomos aqui o que chamaremos de “contexto de comando” - uma variação que antecede o conceito de contexto trazido por Pangaro.

Este conceito consiste em compreender que as interações dos usuários com os sistemas são percebidas como limitadas pois estão definidas pela configuração pré estabelecida deste. Esta definição delimita a capacidade de “resposta-ação” do comando feito pois, mesmo que novas informações sejam inseridas o sistema não está apto a captar esses inputs e transformar em referência base para novas interações posteriores. Ou seja, não há possibilidade de que novas informações modifiquem ou ampliem o contexto a fim de possibilitar interações não pré-programadas. Assim, podemos então presumir que, por mais que esses sistemas tenham capacidade de aprender ao longo do uso esse aprendizado ainda é incapaz de gerar percepções por parte da maioria dos usuários visto que o contexto de aprendizagem do sistema está ligado ao seu código estrutural.

7.2

Linguagem

Quando observamos as definições empregadas por Pangaro (2017, p. 1580) ao conceito de contexto, notamos a sua ligação com o significado de linguagem. É através dele que podemos conceber a linguagem, já que o contexto "prepara o cenário para a linguagem" (Pangaro, 2017, p. 1580). Desta forma, como propomos outro conceito para o significado de contexto para interações com sistemas conversacionais, é necessário avaliar também o conceito de linguagem empregado nesses sistemas.

Pangaro (2017, p. 1580) define a linguagem como "ao menos um significado inicial compartilhado que transmita sentido, para começar uma [...]". Se, pela

definição de Pangaro (2017), contexto e linguagem estão interligados, aqui não se dará de forma diferente. Contudo, como a definição de contexto se diferencia nesta nova proposta, isso acaba inferindo também em diferenças no conceito de linguagem. Isso porque é o contexto que prepara o cenário de interação para a linguagem e o significado de contexto aqui é o de “contexto de comando”, como explicitado na seção anterior.

Assim, inferimos que a linguagem é afetada pela redução da compreensão do contexto. Ou seja, a linguagem compreendida é atrelada às configurações pré-estabelecidas ao sistema, ou seja, ao “contexto de comando”. Ao termos uma interação entre humano-máquina em que a linguagem está limitada à capacidade da máquina de gerar contextos novos para o diálogo todo este sistema conversacional fica restringida à capacidade da máquina de apreender novos contextos a partir de experiências prévias nas conversas. A máquina sendo limitada ao seu código inicial faz com que aprendizados comuns ao ser humano precisem ser reencenados a cada nova interação como se não tivessem sido executados antes.

7.3

De troca para comando

Pangaro (2017, p. 1580) define que o terceiro conceito dentro do modelo que expressa uma conversação é o de troca, definido pelo autor como "engajamento recíproco, em interações baseadas em linguagens que podem resultar em acordo" (Pangaro, 2017, p. 1580). A definição empregada pelo autor descreve a interação entre dois sistemas capazes de tomar decisões, pois são capazes de gerar um engajamento recíproco.

Como observado na seção 7.1, sistemas conversacionais atualmente não são capazes de tomar decisões, o que descaracteriza o conceito de troca nesta proposta, dando lugar ao conceito de comando que, segundo o dicionário é o ato ou efeito de comandar (Michealis, 2020).

Anteriormente neste estudo, por meio da análise de conteúdo das entrevistas realizadas, inferimos que os sistemas ditos como conversacionais atualmente são percebidos por seus usuários como capazes de manter apenas um turno de interação. Ou seja, eles são capazes de ouvir o comando de seus usuários e avaliar se estão

dentro ou não do “contexto e da linguagem de comando”, para que então possam calcular se o comando dado está de acordo com as configurações pré-estabelecidas, realizando o pedido através de uma “resposta-ação”, que por sua vez será capaz de inferir alguma percepção ao usuário, finalizando assim o ciclo de interação.

Nessa situação, é possível evidenciar que não há trocas contínuas entre as partes envolvidas, mas comandos dados por seus usuários, os quais, de acordo com as configurações pré-estabelecidas, são atendidas, finalizando o ciclo de interação.

7.4

Acordo

O conceito de acordo definido por Pangaro (2017, p. 1580) é de "(suficientemente) entendimento compartilhado de conceitos, intenções, valores, que podem levar a [...]". Nesta proposta, também é empregado o conceito de “acordo”. Contudo, ele não é mantido com a essência descrita pelo autor. Neste trabalho, acordo se configura como sendo entendimento suficiente por parte do usuário sobre as pré-configurações do sistema no qual a única tarefa empregada ao sistema é revisar suas pré-configurações na intenção de definir se é capaz ou não de realizar o comando e, reside também, na confiança da credibilidade das informações recebidas.

7.5

De ação ou (trans)ação para resposta-ação

No modelo conversacional desenhado por Pangaro (2017) o quinto e último conceito empregado é o da ação ou (trans)ação que é definido pelo autor como "uma interação coordenada em domínios que não sejam a linguagem, por exemplo, comércio, contratos, dança, jogos." (2017, p. 1580)

Nesta proposta, podemos avaliar que a definição de ação ou (trans)ação não é totalmente contemplada, visto que a resposta do sistema pode pertencer aos domínios da linguagem, sendo uma função evidente na pré-configuração do sistema. Isso é verificado no seguinte trecho - pertencente à unidade de registro [1UR2.3] - retirado da entrevista: *‘O principal uso que eu faço é como alto-falante e pergunto sobre o tempo, a temperatura, se vai chover ou não. Esses são os*

principais uso que eu faço’ (Entrevistado 18). E neste outro - pertencente à unidade de registro [1UR9.1]: *‘Eu acho que quando chegou e eu ficava testando, ficava brincando com ela. Tipo ‘Onde a gente tá, onde eu tô?’ Aí ela fala a cidade.’* (Entrevistado 9). É importante ressaltar que o modelo de Pangaro (2017) cobre parcialmente o conceito de ação ou (tras)ação, já que é possível que através das interações certas ações possam ser tomadas, tais como pedir um uber ou uma refeição no ifood, o que pontuamos aqui, é que os sistemas baseados em interações por voz vão além da definição do autor quando ele desvalida a fala (linguagem) como ação ou (trans)ação.

Considerando que o conceito de “ação ou (trans)ação” não corresponde às necessidades totais desta proposta, indicamos o uso do conceito “resposta-ação”, uma composição feita através das palavras “resposta” e “ação”, ambas dicionarizadas da seguinte maneira: (1) “Resposta: ato ou efeito de responder, utilizando palavras ou gestos; explicação ou esclarecimento de determinada questão (Michaelis, 2020); (2) “ação: ato ou efeito de agir; faculdade ou possibilidade de agir, de executar alguma coisa” (Michaelis, 2020). Sendo assim, nesta pesquisa, definimos o termo “resposta-ação” como: a capacidade de resposta de sistemas baseados em interações por voz utilizando palavras e/ou executando ações.

7.6

Feedback/Percepção

O modelo de Pangaro (2017, p. 1580) simula uma interação baseada em trocas constantes, a fim de se obter um acordo sobre o tema discutido sendo capaz de modificar, pelo menos, uma das partes envolvidas ao final da interação. Construímos ao longo de todo este capítulo a proposta de um modelo que se adeque às necessidades específicas da interação entre o humano e os sistemas baseados em interações por voz, baseando-nos em Pangaro (2017) o qual, por sua vez, utiliza-se do modelo canônico de *feedback-loop* proposto por Kroemer e Grandjean (2007, p. 125).

Observando o modelo proposto por Kroemer e Grandjean (2007), consideramos que a última etapa que completa o ciclo é o *feedback*. É a percepção que permite que os usuários interpretem e tomem sua próxima decisão dando sequência ao fluxo de interação, conforme esta proposta exemplifica a seguir.

7.7

Revisão do modelo conversacional

A revisão do modelo conversacional proposto nesta seção é construída com base em todas as inferências percebidas ao longo deste trabalho e descritas no capítulo anterior. Possui seis conceitos chaves: (1) Contexto de comando - contexto limitado às pré-configurações do sistema capaz de preparar o cenário para [...]; (2) Linguagem - ao menos um significado inicial pré-configurado ao sistema que transmita sentido, para começar um [...]; (3) Comando - ato ou efeito de comandar em interações baseadas em linguagens que podem resultar em [...]; (4) Acordo - entendimento suficiente por parte do usuário sobre as pré-configurações do sistema que podem levar a [...]; (5) Resposta-ação - capacidade de resposta de sistemas baseados em interação por voz utilizando palavras e/ou executando ações, resultando em [...]; (6) *Feedback*/Percepção - recebimento de dados capaz de inferir em interpretações e tomadas de decisões.

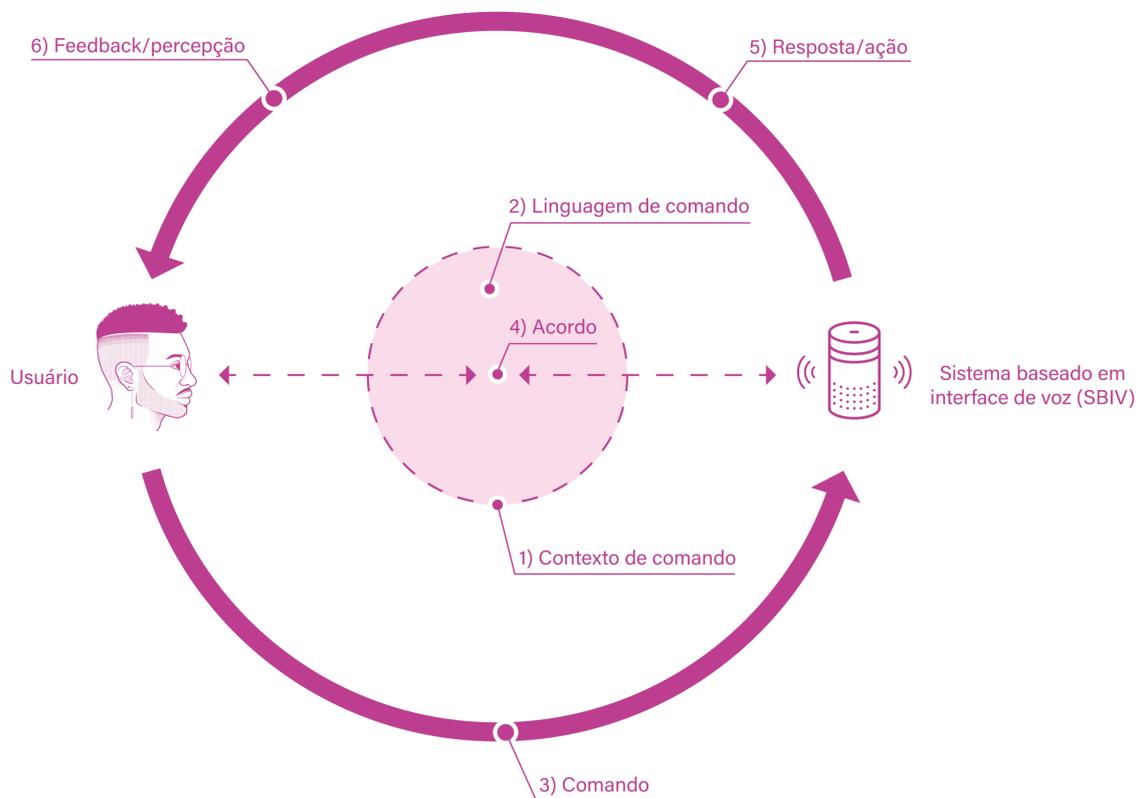


Figura 14 - Revisão do modelo conversacional. Fonte: autor (2020)

Partindo da revisão ao modelo conversacional proposta nesta pesquisa, é possível afirmar que ainda é inconcebível inferir uma interação conversacional plena aos sistemas atuais baseados em interação de voz. As trocas não contínuas precisam retornar ao seu ponto de origem – o interlocutor humano – para que o modelo conversacional pelo uso de voz ganhe o aspecto de interação linguística pela ótica humana mesmo que seu ciclo retorne à uma origem contextual prévia da máquina. Ou seja, o sistema retorna ao início para nova troca sem o aprendizado contextual novo inerente a conversação.

8

Conclusão e desdobramentos futuros

A realização desta pesquisa permite trazer para o campo do *design* - inserido na linha de pesquisa Ergonomia e Usabilidade e Interação Humano-Computador - um aprofundamento no que diz respeito à fala como forma de interação humano-máquina, visto que há muitas dúvidas que circundam este modelo de interação.

Esta pesquisa, além de fomentar discussões e aprofundamentos sobre o tema proposto, gera valor social no que respeita à produção de material acadêmico relacionado à linha de pesquisa desenvolvida por pesquisadores negros, visto que o universo acadêmico no Brasil é majoritariamente branco. Vale destacar, portanto, dados de estudo proposto pelo IBGE (2018), o qual nos informa que, dentre os jovens de 18 a 24 anos, apenas 55,6% são compostos jovens negros que chegam ao ensino superior, em contraste com os 78,8% de jovens brancos.

Muito além de validar ou não a hipótese proposta nesta pesquisa, é necessário entendermos que contribuições científicas como essa muitas vezes são limitadas ao que se entende por elite intelectual, o que marginaliza a maior parte da população brasileira, ou seja, o povo preto. A ruptura com o que entendemos hoje ser ainda a elite intelectual é necessária, portanto, para que cada vez mais a produção de conhecimento científico se espalhe, tornando a academia mais democrática.

Compreendemos que é necessário que a academia deixe suas raízes racistas de lado e incorpore cada vez mais pensadoras e pensadores pretos no ambiente científico. Por isso, não basta não ser racista. Efetivamente, precisamos ser também antirracistas e promover a inclusão do povo preto em todos os âmbitos do conhecimento. Postas essas considerações, damos continuidade à conclusão desta pesquisa a seguir.

8.1

Percurso percorrido ao longo da dissertação

Para concluirmos este estudo, é válido retomar os caminhos que guiaram e fundamentaram esta pesquisa, a qual teve seu início a partir de ocasionais observações acerca da comunicação das pessoas envolvendo sistemas baseados em interação por voz. Nisso constatamos que é comum que tais interações sejam interrompidas ou não concluídas com êxito. Essa percepção é tão comum, que observamos representações desses "erros" em grandes produções cinematográficas, como retrata o filme *Nós* (2019), do diretor Jordan Peele.

Por tudo isso, este estudo apresenta-se da seguinte forma: o capítulo 1, "Introdução", abrangeu: (1) delimitação do universo da pesquisa; (2) os objetivos, o percurso metodológico e a justificativa da investigação.

No capítulo 2, "Sociedade e Comunicação", foram abordados: (1) caracterização da comunicação, a partir da cronologia humana; (2) definição de comunidade primitiva; (3) definição de sociedade tradicional; (4) definição de sociedade complexa; (5) a democratização da microinformática; e (6), a disseminação da *internet* como ferramenta e técnica.

Já no capítulo 3, "Interface: uma breve descrição", abordamos a definição do conceito de interface na literatura pertinente.

Em "*Design* de interação: sistemas conversacionais", capítulo 4, apresentamos: (1) definição do termo *design* de interação; (2) definição de interação; (3) como funcionam as atividades mentais humanas, quando em interação com as máquinas; (4) conceito de interação sonora; (5) conceito e definição de interação por voz e (6) modelo de interação baseado em conversação.

No capítulo 5, "Delineamento da pesquisa", apresentamos: (1) o delineamento da estrutura de pesquisa; (2) métodos e técnicas; (3) a entrevista como instrumento de coleta de dados. As informações obtidas por meio da técnica de entrevista contribuíram com dois aspectos: com a elaboração da proposta de revisão do modelo conversacional, e com as considerações em torno das experiências no uso de sistemas baseados em interação por voz.

Já em "Análise de conteúdo como ferramenta analítica", no capítulo 6, a estrutura é a seguinte: (1) descrição do processo adotado; (2) unidades de registro, unidades de contexto e categorias extraídas da análise de conteúdo das respostas obtidas através das entrevistas realizadas; (3) análise de conteúdo e inferências. Ressaltamos que as informações obtidas através da análise de conteúdo foram

cruciais para a configuração e elaboração da proposta de modelo interacional proposto nesta pesquisa.

No capítulo 7, intitulado “Discussão”, definimos o seguinte: (1) as diferenças e as relações do modelo proposto com o modelo conversacional apresentado ao final do capítulo 4; (2) o conceito de “contexto de comando”; (3) o conceito de linguagem empregado na proposta de modelo; (4) o conceito de comando; (5) o conceito de “acordo”; (6) o conceito “resposta-ação”; (7) o conceito de “*feedback*/percepção”; (8) a ilustração do modelo e suas considerações.

A partir dos resultados obtidos ao longo de toda a dissertação, estruturamos o capítulo 8, “Conclusão e desdobramentos futuros”, com os seguintes tópicos: (1) percurso percorrido ao longo da dissertação, (2) conclusão e (3) desdobramentos futuros.

8.2

Conclusão

A partir dos dos objetivos alcançados através da revisão bibliográfica, entrevista, análise de conteúdo e dos resultados dispostos, concluímos que a tecnologia em seu estado atual não é capaz de contemplar os conceitos que definem o modelo conversacional apresentado nesta dissertação, validando a hipótese proposta.

O modelo de Pangaro (2017) mostra a incapacidade desses sistemas no reconhecimento e compreensão dos mais diversos contextos nos quais os usuários possam estar inseridos, limitando a compreensão da linguagem, dos comandos e das ações as suas pré-configurações.

Isso implica numa linha de aprendizado - não sob a ótica do sistema, mas sobre a ótica do usuário. Esse, baseando-se em tentativas e erros, deve aprender e entender a capacidade de seu sistema, criando assim uma compreensão sobre os seus limites contextuais em uso, o que infere nas limitações já observadas e registradas neste estudo e que nos levou a constatar que o contexto explorado é o do sistema e não o do usuário.

Por mais que a fala seja abordada como a linguagem mais natural ao ser humano, ela é natural apenas ao ser humano que, graças ao seu complexo sistema de atividades mentais, é capaz de aprender, reaprender e modificar as suas várias

formas de linguagem ao longo dos anos e dos diferentes contextos. Isso é algo que a máquina ainda não consegue replicar e, por isso, tal fato acaba interferindo diretamente na maneira pela qual se configura a experiência do usuário, que espera uma troca tão próxima quanto aquelas obtidas em conversas com outros seres humanos. Por isso, concluímos que a relação que existe entre um ser humano e um sistema, baseado em interação por voz, não se dá através de conversas, mas por meio de comandos, conforme demonstra a discussão dos resultados, transformando este apontamento na principal reflexão desta pesquisa.

A revisão do modelo conversacional de Pangaro (2017) como conclusão dessa pesquisa destaca que há inferências na relação humano-SBIV (Sistema Baseado em Interface de Voz), de forma a se diferenciar em alguns aspectos e conceitos e de se aproximar em outros. Quando verificamos o modelo conversacional proposto pelo autor, observamos estes conceitos empregados num modelo de *feedback-looping*: (1) Contexto - um momento, situação, lugar e/ou história compartilhada que prepara o cenário para [...]; (2) Linguagem - ao menos um significado inicial compartilhado que transmita sentido, para começar uma [...]; (3) Troca - engajamento recíproco, em interações baseadas em linguagens que podem resultar em [...]; (4) Acordo - (suficientemente) entendimento compartilhado de conceitos, intenções, valores, que podem levar a [...]; (5) Ação ou (trans)ação - uma interação coordenada em domínios que não sejam a linguagem, por exemplo: comércio, contratos, dança, jogos.

Ao observarmos o recorte dos pares dentro da interação - o homem e o sistema baseado em interação por voz -, verificamos que se evidenciam necessidades específicas nesse modelo de interação vigente aos pares.

A revisão do modelo também é baseada no conceito de *feedback-looping* e mantém proximidades com o modelo conversacional, como é possível observar quando analisamos os conceitos empregados em sua construção: (1) Contexto de comando - contexto limitado as pré-configurações do sistema capaz de preparar o cenário para [...]; (2) Linguagem - ao menos um significado inicial pré configurado ao sistema que transmita sentido, para começar um [...]; (3) Comando - ato ou efeito de comandar em interações baseadas em linguagens que podem resultar em [...]; (4) Acordo - entendimento suficiente por parte do usuário sobre as pré configurações do sistema que podem levar a [...]; (5) Resposta-ação - capacidade de resposta de sistemas baseados em interação por voz utilizando palavras e/ou executando ações,

resultando em [...]; (6) *Feedback*/Percepção - recebimento de dados capaz de inferir em interpretações e tomadas de decisões.

8.3

Desdobramentos futuros

Para olharmos para frente, devemos nos atentar ao passado e, com isso em mente, traçamos um olhar retrospectivo capaz de evidenciar o que este estudo contempla: (Ao considerarmos os possíveis desdobramentos desta pesquisa, lançamos um olhar retrospectivo nesta trajetória, que evidencia as seguintes contribuições deste estudo): a importância do cenário que contextualiza o uso de sistemas baseados em interação por voz - Assistentes Virtuais Inteligentes e assistentes de casa (alto-falantes inteligentes); o levantamento teórico sobre a comunicação humana e suas evoluções; os fundamentos ergonômicos que são os pilares que sustentam toda essa pesquisa; a técnica de entrevista e a análise de conteúdo, como estratégia para obtenção de dados dos usuários; a proposta de um novo modelo interativo; e, por fim, as considerações finais em torno da experiência no uso de modelos baseados em interação por voz.

Esta pesquisa é orientada por abordagens metodológicas apoiadas na conduta científica. Mesmo assim, é importante ressaltarmos que a abordagem assumida durante o estudo, os recortes dos fundamentos teóricos e a construção das entrevistas refletem, sem sombras de dúvidas, as nossas crenças e perspectivas. Ainda, salientamos que há pouquíssima produção de material científico em língua portuguesa. Por isso, esperamos ter contribuído com esta pesquisa, da seguinte maneira: (1) disponibilizando informações acerca da comunicação entre o humano e a máquina através da voz; (2) descrevendo o modelo interacional entre o humano e sistemas baseados em interface de voz; e, por fim, (3) dispondo de material teórico e analítico sobre o tema desta pesquisa, em língua portuguesa.

Ressaltamos que, mesmo que o objetivo dessa pesquisa tenha sido alcançado, este estudo pode ser aprofundado, a fim de alcançar novos desdobramentos, como: (1) validar de maneira científica a revisão feita nesta pesquisa do modelo de Pangaro ; (2) abordar os fundamentos teóricos de cada conceito empregado nesta revisão, como por exemplo entender o por que de os usuários não percebem o aprendizado do sistema em seu uso contínuo; (3) promover a utilização do modelo

interacional revisado na aplicação de novos sistemas baseados em interação por voz; (4) compreender as implicações do turno único no design de interação e no UX; (5) utilizar o conteúdo gerado nessa pesquisa como alicerce para a comparação entre *chatbots* e outras instâncias da I.A.; (6) e investigar a fundo como seriam as possibilidades e viabilidades para transformar comandos em conversas.

9

Referências bibliográficas

#207 DESIGNING VOICE INTERFACES WITH BEN SAUER. [Locução de]: Per Axbom e James Royal-Lawson. UX Podcast, 23 Mar. 2019. *Podcast*. Disponível em: <<https://open.spotify.com/episode/0fG6PzHumrSMzzJU9UrrkX>>. Acesso em: 20 Jun. 2019.

2019 VOICE REPORT: consumer adoption of voice technology and digital assistants. **Microsoft**. Disponível em: <<https://about.ads.microsoft.com/en-us/insights/2019-voice-report>>.. Acesso em: 22 mar. 2020.

AÇÃO. In: **DICIONÁRIO** Michaelis On Line. São Paulo: Editora Melhoramentos, 2020. Disponível em <https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/comando/>. Acesso 01 ago. 2020

ALGARRA, L. **Teoria da conversação**: Uma amostra do trabalho de Gordon Pask revisitado por Paul Pangaro e adaptado para o português por Luiz Algarra. 2012. 17 slides. Disponível em: <<https://pt.slideshare.net/lalgarra/teoria-das-conversaes-g-pask>>. Acesso em: 3 dez. 2019.

BARBOSA, S. D. J.; SILVA, B. S. da. **Interação humano-computador**. Rio de Janeiro: Elsevier, 2010. 384 p. Disponível em: <<https://docplayer.com.br/63299367-Interacao-humano-computador.html>>. Acesso em: 26 out. 2019.

BARDIN, L. **Análise de conteúdo**. Brasil: Almedina, 2016. p. 53-210.

BRAZIL, E; FERNSTRÖM, M. Auditory Icons. In: HERMANN,T; HUNT, A.; NEUHOFF, J. G. (Eds.).**The Sonification Handbook**. Berlin, Germany: Logos Publishing House: 2011. p. 325 – 338.

CASTELLS, M. **A sociedade em Rede**. São Paulo: Paz e Terra, 1999. v.1, 4.ed., 630 p.

CARLOMAGNO, Márcio C; ROCHA, Leonardo Caetano da. COMO CRIAR E CLASSIFICAR CATEGORIAS PARA FAZER ANÁLISE DE CONTEÚDO: uma questão metodológica. Revista Eletrônica de Ciência Política, [S.L.], v. 7, n. 1, p. 173-188, 18 jul. 2016. Universidade Federal do Parana. <http://dx.doi.org/10.5380/recp.v7i1.45771>. Disponível em: https://www.researchgate.net/publication/305416244_COMO_CRIAR_E_CLASSIFICAR_CATEGORIAS_PARA_FAZER_ANALISE_DE_CONTEUDO_UMA_QUESTAO_METODOLOGICA. Acesso em: 10 maio 2019.

COMANDO. In: **DICIONÁRIO** Michaelis On Line. São Paulo: Editora Melhoramentos, 2020. Disponível em <https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/comando/>. Acesso 01 ago. 2020

DIRINGIR, David. A history of the alphabet. 3ed. London: Utchinson, 1968, p. 14.

DUNN, J. Virtual assistants like Siri and Alexa look poised to explode. Disponível em: <businessinsider.com>

DUBBERLY, H; PANGARO, P. Cybernetics and design: conversations for action. In: FISHER, T.; HERR, C.M. Desing Cybernetics: Navigating and design. **Springer**. Nova York: 2019. p.85-99.

_____; HAQUE, U. What is interaction? Are there different types?. **Interactions**. New York, v.16, n.1, p. 69-75, 2009. Disponível em: <http://www.dubberly.com/wp-content/uploads/2009/01/ddo_article_whatinteraction.pdf>. Acesso em: 26 out. 2019.

ENGRAMA. In: **DICIONÁRIO** Michaelis On Line. São Paulo: Editora Melhoramentos, 2020. Disponível em <https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/comando/>. Acesso 22 ago. 2020

FONTANELLA, B. J. B.; RICAS, J.; TURATO, E. R. Amostragem por saturação em pesquisas qualitativas em saúde: contribuições teóricas. **Cadernos de Saúde Pública**, v. 24, n. 1, p. 17-27, jan. 2008. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2008000100003&lng=en&nrm=iso>. Acesso em: 20 de jun. 2020.

KROEMER, Karl H.; GRANDJEAN, E. **Manual de Ergonomia: Adaptando o Trabalho ao Homem**. Bookman, 2007. Edição do Kindle.

HALL, E. **Conversational design**. Nova York: A Book Apart, 2018. 420 p.

HOHLFELDT, A.; MARTINO, L. C.; FRANÇA, V.V. (orgs.). **Teorias da Comunicação**. Petrópolis: Editora Vozes, 2008. 337 p.

HUGO, Marcel. **Uma interface de reconhecimento de voz para o sistema de gerenciamento de central de informação de frete**. 1995. 60 f. Dissertação (Mestrado) - Curso de Engenharia de Produção, Engenharia, Universidade Federal de Santa Catarina, Santa Catarina, 1995.

HUNT, A.; HERMANN, T. Interactive Sonification. In: HERMANN,T.; HUNT, A.; NEUHOFF, J.G. (Orgs.).**The Sonification Handbook**. Berlin, Germany: Logos Verlag, 2011. p. 273-298

IBGE. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Desigualdades sociais por cor ou raça no Brasil**. Rio de Janeiro: IBGE, 2019. Disponível em https://biblioteca.ibge.gov.br/visualizacao/livros/liv101681_informativo.pdf. Acesso em jul. 2020.

KEYNES, M. **Strategy Analytics**: new record for smart speakers as global sales reached 146.9 million in 2019. Businesswire – A Berkshire Hathaway Company. 2020. Disponível em: <<https://www.businesswire.com/news/home/20200213005737/en/Strategy-Analytics-New-Record-Smart-Speakers-Global>>. Acesso em: 20 mar. 2020.

LEMOS, A. **Cibercultura**: Tecnologia e vida social contemporânea. Porto Alegre: Sulina, 2010. 296 p.

LIMA, G. A. B. **A transmissão do conhecimento através do tempo**: da tradição oral ao hipertexto. Revista Interamericana de Bibliotecología, v.30, n.2, p. 275-285, 2007. Disponível em: <http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-09762007000200013&lng=en&nrm=iso>. Acesso em: 18 maio 2019.

MARTINO, L. **A Revolução Mediática**: a comunicação na Era da simulação tecnológica. Razón y palabra. n. 49, 2005. Disponível em: <https://www.researchgate.net/publication/28131940_A_Revolucao_Mediatica_a_comunicacao_na_Era_da_simulacao_tecnologica>. Acesso em: 14 maio 2019.

MCGOOKIN, D.; BREWSTER, S. E. Earcons. In: HERMANN, T.; HUNT, A.; NEUHOFF, J. G. (org.). **The Sonification Handbook**. Berlin, Germany: Logos Verlag, 2011. p. 339–361

MORAES, Roque. **Análise de conteúdo**. Revista Educação, Porto Alegre, v. 22, n. 37, p. 7-32, 1999.

NEUHOFF, J. G. Perception, Cognition and Action in Auditory Displays. In: HERMANN, T.; HUNT, A.; NEUHOFF, J. G. (orgs.). **The Sonification Handbook**. Berlin, Germany: Logos Verlag, 2011. p. 273-298.

NORMAN, Donald A., personal correspondence, 31 de outubro, 2008.

NIELSEN LAUNCHES NEW MEDIATECH TRENDER SURVEY TO UNCOVER CONSUMER SENTIMENT ON EMERGING TECHNOLOGIES. Nielsen. Disponível em <<https://www.nielsen.com/us/en/press-releases/2018/nielsen-launches-new-mediatech-trender-survey-to-uncover-consumer-sentiment-on-emerging-technology/>>. Acesso em: 30 abr. 2020.

NÓS. Direção de Jordan Peele. Produção de Jason Blum, Jordan Peele, Sean McKittrick, Ian Cooper. Blumhouse Productions, Monkeypaw Productions, QC Entertainment, 2019. Longa metragem (116). Disponível em <https://www.telecineplay.com.br/filme/Nos_15554>. Acesso em: 28 jun. 2020.

OLIVEIRA, A. S. **Smartphones e trabalho imaterial**: uma etnografia virtual sobre sujeitos usuários de dispositivos móveis convergentes. Dissertação (Mestrado em Comunicação), UFRGS. Porto Alegre, 2007. 136p. Disponível em: <<https://lume.ufrgs.br/handle/10183/12142>>. Acesso em: 27 maio 2019.

PANGARO, P. Questions for conversation theory or conversation theory in one hour. **Kybernetes**, Detroit, v. 46, n. 9, p.1578-1587, Disponível em: <

<https://www.emerald.com/insight/publication/issn/0368-492X>> Acesso em 2 out. 2019.

PASK, G. **Conversation cognition and learning: a cybernetic theory and methodology**. Amsterdam: Elsevier, 1975. 570 p.

PEARL, C. **Designing voice user interfaces**. Sebastopol: O'reilly, 2016. 278 p.

PRODANOV, Cleber Cristiano; FREITAS, E. C. D. Metodologia do trabalho científico: : Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico. 2. ed. Novo Hamburgo - Rio Grande do Sul - Brasil: Universidade Feevale, 2013. p. 11-276.

RESPOSTA. In: **DICIONÁRIO** Michaelis On Line. São Paulo: Editora Melhoramentos, 2020. Disponível em <https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/comando/>. Acesso 01 ago. 2020

RODRIGUES, M. H. C. Gutemberg e o letramento do ocidente. **Revista Educação e Linguagens**, v. 1, n. 1, ago./dez. 2012.

SAFFER, D. **Designing for interaction: Creating Innovative Applications and Devices**. 2.ed. Berkeley: New Riders, 2010. 241 p.

SERAFIN, S. Sonic Interaction Design. In: HERMANN,T; HUNT, A; **The Sonification Handbook**. Berlin, Germany: Logos Verlag, 2011.p. 88-107

SHAHANI, A. (2016). **Voice Recognition Software Finally Beats Humans At Typing, Study Finds**. Disponível em <http://npr.org/>. Acesso em: 11 jul. 2020.

THRO, Ellen. The artificial intelligence dictionary. The Lance. A. Leventhal Microtrend Series. San Marcos: Microtrend, 1991.

VELOSO, D. **Inovação no Design de Interação: Design livre para modelos sustentáveis**. 2011. 40 p. Dissertação (Mestrado) - Curso de Pós-graduação em Design de Interação, Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 2011.

VERPLANK, Bill, Interaction Design Sketchbook, February, 2001.

WARGO, M. J. (1967). Human operator response speed, frequency and flexibility: a review and analysis. *Human Factors*, 9, 221-38.

WARREN, J. A. The First Speech Recognition System. **Asta Speaks**, 2014. Disponível em: <<https://astaspeaks.wordpress.com/2014/10/13/audrey-the-first-speech-recognition-system/>>. Acesso em: 26 maio 2019.

A Termo de Consentimento Livre e Esclarecido

Título da pesquisa: Interfaces por voz: princípios e conceitos para realização de tarefas eficientes

Professora orientadora: Dr^a. Maria Manuela Rupp Quaresma

Mestrando: Lucas José da Silva Santos

Você está sendo convidado a ser um participante/voluntário desta entrevista.

Objetivo:

Essa entrevista faz parte de uma pesquisa em desenvolvimento no curso de mestrado em Design, na linha de pesquisa: Ergonomia, Usabilidade e Interação Humano-Computador. O objetivo dessa entrevista é colher insumos para uma melhor concepção por parte do pesquisador sobre o universo que circunda interfaces do usuário por voz.

Justificativa

Com essa pesquisa pretende-se analisar a percepção do entrevistado em relação a eficiência das interações realizadas através de interfaces do usuário por voz. A entrevista irá gerar insumos importantes para a análise e proposições deste estudo.

Procedimentos

Se você decidir fazer essa parte dessa pesquisa, o pesquisador efetuará perguntas em formato de conversa, se baseando em tópicos previamente definidos. A entrevista será através de meio digital, através de chamada de voz, não necessitando da presença física de quaisquer participantes. É importante pontuar que toda a entrevista estará sendo gravada para posterior decupagem e análise, respeitando o período mínimo de guarda de cinco anos pelo entrevistador. A entrevista tem duração aproximada de 50 minutos, tendo o entrevistado liberdade total para responder as perguntas realizadas.

Risco

Não há riscos previstos.

Benefícios

Sua participação na entrevista contribuirá para uma melhor compreensão da tecnologia em pauta, podendo transformar a forma com a qual a sociedade interage e se comunica através da fala com assistentes virtuais e assistentes de casa.

Compensação

Não há nenhuma remuneração por sua participação nesta entrevista.

Informações coletadas

A entrevista será gravada em áudio e o entrevistador tomará notas, além de utilizar o arquivo em registro para futura decupagem e análise.

Sigilo

Para proteger sua imagem, o seu nome não aparecerá em nenhuma publicação. Você receberá um pseudônimo (nome falso): usuário 01, por exemplo.

Autorização para uso e registro de áudio

Ao assinar este termo, você autoriza o uso de todo áudio registrado, dentre: suas declarações e sua voz para finalidades acadêmica – artigos acadêmicos, aulas, *papers*, sites, apresentações em simpósios e/ou congressos científicos relacionados ao tema.

Custo para você

Não há custos.

Direitos dos participantes

Sua participação nesta entrevista é voluntária e, portanto, não há nenhuma obrigação de participar. Se você necessitar de uma pausa a qualquer momento durante a entrevista, por favor, avise. Você tem o direito de mudar de ideia e interromper a entrevista a qualquer momento, sem apresentar motivos e sem qualquer penalização. Qualquer nova informação que possa fazê-lo mudar de ideia

sobre sua participação na pesquisa será fornecida a você. Você receberá uma cópia deste documento de consentimento. Não há renúncia a qualquer de seus direitos legais ao assinar ou concordar com este termo.

Perguntas

Você poderá intervir e questionar sempre que achar necessário ou tiver alguma dúvida.

Se tiver alguma dúvida sobre esta pesquisa, você também pode contatar o mestrando Lucas José da Silva Santos pelo telefone (21) 97991-7419 ou lucasabdack@gmail.com, ou a professora orientadora responsável Prof. Dr^a. Maria Manuela Rupp Quaresma, (21) [3527 1005](tel:35271005) ou

Este documento resultará em duas vias, uma que ficará em posse do entrevistador e outra em posse do entrevistado.

Data: ___/___/___

Assinatura do aluno:

Assinatura da professora:

Seu nome (por extenso):

Assinatura:
