Acta Scientiarum
http://periodicos.uem.br/ojs
ISSN on-line: 1983-4683
Doi: 10.4025/actascilangcult.v44i1.60071

A voz do estereótipo nos assistentes digitais

Joaquim Braga

Faculdade de Letras, Instituto de Estudos Filosóficos, Universidade de Coimbra, Largo da Porta Férrea, 3004-530, Coimbra, Portugal. E-mail: bragajoaquim77@gmail.com

RESUMO. Os processos de reprodução e sintetização da voz humana nos dispositivos de assistência digital não obedecem, somente, a mecanismos puramente tecnológicos. Neles também se evidenciam quer formas de conceber a linguagem segundo os preceitos da informação quer formas de actualizar e perpetuar certos estereótipos sociais, particularmente os de género. É com o propósito de reflectir sobre a articulação de ambas que se justificam os conteúdos expostos no presente artigo, assim como a principal distinção conceptual, que, nele, é desenvolvida, entre 'a voz que ouve' e 'a voz que é ouvida'. São duas as teses que norteiam esse propósito. A primeira sugere que a padronização dos assistentes digitais segundo critérios de discriminação de género é, ainda, nos nossos tempos, uma das várias manifestações do amplo fenómeno da divisão sexual do trabalho. A segunda tese, por sua vez, abrange o fenómeno tecnológico da chamada "convergência digital" e diz respeito aos modos de como os estereótipos auditivos, baseados na imagem de mãe-cuidadora, complementam e reforçam os estereótipos visuais. Para que a articulação teórica das duas teses seja profícua, é necessário, antes de tudo, desmistificar a suposta interacção verbal entre máquina e utilizador, mostrando – ao contrário do que se infere do imaginário tecnológico contemporâneo – a impossibilidade dialógica de ambos.

Palavras-chave: género; incorporação; informação; linguagem; tecnologia.

The voice of stereotype in digital assistants

ABSTRACT. In digital assistance devices, human voice reproduction and synthesis processes do not just are ruled by purely technological mechanisms. They also show both ways of conceiving language according to information patterns and ways of updating and perpetuating certain social stereotypes, particularly those of gender. It is with the aim of inquiring into the articulation of both that the contents exposed in this paper are justified, as well as the main conceptual distinction, which is developed through it, between 'the voice that hears' and 'the voice that is heard'. Two significant theses guide such purpose. The first suggests that the standardization of digital assistants according to gender discrimination criteria is still, in our times, one of several expressions of the broad phenomenon of the sexual division of labour. The second thesis, in turn, covers the technological phenomenon of the so-called "digital convergence" and concerns how auditory stereotypes, based on the image of the mother-caregiver, complement and reinforce visual stereotypes. For the theoretical articulation of these two theses to be feasible, it is necessary, above all, to demystify the supposed verbal interaction between machine and user, showing – contrary to what is inferred from the contemporary technological imaginary – the dialogical impossibility of both.

Keywords: gender; embodiment; information; language; technology.

Received on July 13, 2021.
Accepted on August 6, 2021.

Introdução

Com o desenvolvimento e incremento dos meios tecnológicos digitais nas actividades humanas quotidianas, tem havido, gradualmente, uma transformação das formas como usamos e concebemos a linguagem. Um dos casos mais paradigmáticos das mudanças que as novas tecnologias operam na esfera linguística diz respeito à função remissiva que a hipertextualidade digital veio acrescentar aos símbolos da linguagem escrita. A palavra, além de cumprir o seu poder de significação, empresta os seus próprios caracteres ao dispositivo de remissão que permite sair de um texto para entrar num outro texto, semelhante ou díspar. O mesmo é dizer: os símbolos discursivos escritos deixam-se, simultaneamente, sobrepor pelos sinais indicativos de que há mais informação fora do texto da sua inscrição.

Idêntico fenómeno pode ser vislumbrado no domínio vocal da linguagem. Também, aí, se tem assistido a uma crescente utilização tecnológica da voz humana, tanto para activar determinadas operações computorizadas quanto para veicular informações que essas operações permitem. Tal como sucede com a palavra no hipertexto, é acrescentada à voz uma função indexical, por meio da qual se torna duplamente possível sinalizar e percepcionar o que se pretende ouvir. Em rigor, seria um equívoco implicar, aqui, o verbo 'falar', uma vez que não há uma relação dialógica entre máquina e utilizador, mas, sim, um processo de troca de informações suportado vocalmente. Quando ocorrem, as verbalizações do utilizador são instruções de comando dadas à máquina, da mesma maneira que os conteúdos transmitidos pela voz digitalizada correspondem a essas instruções.

A voz não pode ser reduzida a um fenómeno neurofisiológico constituído pela mera produção e emissão de sons. Quando tal se verifica é porque ela regride ao nível da fonação e, deixando de ser um fenómeno relacional, acusa a suspensão da projecção do corpo nos actos de falar e ouvir. A voz é articulada pela modulação da alteridade, ainda muito antes de poder articular sons, fonemas, notas musicais. Sem essa articulação dada pela voz do outro, jamais ela incorporaria o impulso de comunicação, de querer fazer ouvir e de saber ouvir, quando se remete ao silêncio para que o outro possa também falar. Se a voz incorpora e projecta o ouvir e o falar do outro, o que sucede quando lhe é imposto o duplo processo tecnológico de reprodução e sintetização, como no caso dos assistentes digitais?

Ora, embora os dispositivos de assistência digital acusem, irremediavelmente, os limites tecnológicos da reprodução da voz humana e da natureza da interface 'utilizador-máquina', eles tendem a ser concebidos segundo critérios de *design* antropomórficos. Estes visam anular os efeitos da mera obtenção de informação e insinuar uma pretensa interacção verbal entre máquina e utilizador, pela qual as instruções se transmutam em 'perguntas' e os conteúdos disponibilizados, em 'respostas'. Porém, se não preexistem requisitos de género para quem faz as 'perguntas', o mesmo não acontece com quem dá as 'respostas' – o uso da voz feminina é, neste caso, o principal estereótipo que anima a configuração sónica dos assistentes digitais.

Pelo já exposto, torna-se, pois, impreterível analisar os pontos de articulação do fenómeno da voz mediada tecnologicamente com os efeitos latentes dos estereótipos sociais na constituição dos próprios processos de mediação. Como a esse respeito se verá, a inabalável latência do estereótipo remete a uma série de transformações que ocorrem na intersecção da linguagem com a tecnologia e por meio da qual se igualmente reformulam as dimensões somáticas envolvidas no falar e no ouvir humanos.

Linguagem, tecnologia, informação

Na terceira década do século vinte, Homer Dudley, um dos maiores precursores na criação de dispositivos de sintetização da voz, consubstanciou, teoricamente, as suas investigações, recorrendo a nexos analógicos entre a fala humana e as tecnologias de telecomunicação. Nas suas palavras, "[...] o sistema de transmissão de rádio moderno é, somente, um análogo eléctrico do sistema de transmissão acústica do homem fornecido pela natureza"[1] (Dudley, 1940, p. 495). A analogia permitiu-lhe, igualmente, afirmar que o falar "[...] consiste no enviar, por uma mente, e no receber, por outra mente, uma sucessão de símbolos fonéticos com algum conteúdo emocional associado"[2] (Dudley, 1940, p. 495). Ainda que, de forma tímida, acrescente aos conteúdos da informação uma dimensão afectiva, para Dudley, como para tantos autores que subscrevem as teses da teoria da informação, o que se gera entre o cérebro de quem fala e o cérebro de quem ouve é um "[...] fluxo de som audível [...]"[3] que transporta uma mensagem passível de descodificação (Dudley, 1940, p. 495). Este intercâmbio sonoro entre um emissor e um receptor faz jus ao espírito científico da teoria matemática da comunicação, popularizada, na quarta década do século passado, por Shannon e Weaver (1949). Tal como estes dois autores, também Dudley encara as exteriorizações discursivas sob uma moldura quantitativa, pela qual é possível aferir a informação que cada mensagem transporta.

Com efeito, as tentativas de recriação artificial da voz e fala humanas repousam sobre um fundo tecnocientífico que é extensível ao plano teórico de concepção da própria linguagem. Nesse sentido, podemos distinguir, aqui, três momentos reflexivos complementares que traçam o percurso paralelo da análise da linguagem e do desenvolvimento de dispositivos de reprodução e sintetização da voz humana.

O 'primeiro momento' remete aos famosos engenhos criados por Wolfgang Ritter von Kempelen, dos quais Dudley retirou grandes ensinamentos. No século dezoito, com a sua 'máquina falante' (*Sprechmachine*),

[1] "[...] the modern radio broadcast system is but an electrical analogue of man's acoustic broadcast system supplied by nature".
[2] "[...] consists in a sending by one mind and the receiving by another of a succession of phonetic symbols with some emotional content added".
[3] "[...] an audible sound stream [...]".

intentou von Kempelen não só imitar a voz humana e os processos fisiológicos que a suportam, como, também, aumentar as possibilidades do falar. O processo de criação do dispositivo de von Kempelen baseia-se, essencialmente, numa correspondência mimética dos órgãos vocais com os instrumentos musicais, bastando, para tal, como ele descreve, equiparar o pulmão a um "[…] fole […]", a glote a um "[…] cachimbo de junco […]" e a boca a um "[…] sino de oboé em forma de funil" (von Kempelen, 1791, p. 398). Todavia, logo no início do livro em que apresenta as detalhadas formulações sobre o seu engenho, von Kempelen descreve a linguagem como a capacidade de, por meio de signos, transmitir "[…] sentimentos ou pensamentos" (von Kempelen, 1791, p. 1). A diferença entre a linguagem escrita e a reprodução mecânica da linguagem falada reside, segundo o inventor, apenas na especificidade sensorial dos signos utilizados, ou, como ele reitera, "[…] se se pode inventar uma linguagem baseada em sinais escritos para os olhos, não há razão para acreditar que não se possa inventar uma linguagem baseada em sons para os ouvidos"[4] (von Kempelen, 1791, p. 18).

Trata-se, pois, com a máquina falante, de reproduzir um 'órgão sem corpo' – neste caso, o vocal. Contudo, para que tal seja concebível, a ideia de linguagem é submetida a uma redução instrumental: o falar tem uma função de transmissão (de sentimentos e pensamentos), suportada pelo órgão vocal. Reproduzir, mecanicamente, os sons das letras, sílabas e palavras, implica, assim, fazer da transmissão de informação a principal função da linguagem e, por via dela, onerar o próprio falar com elementos que não podem ser abstraídos e isolados dos contextos de interacção, especialmente os que se encontram acoplados com a percepção dos interlocutores, como, por exemplo, os elementos gestuais.

Mais do que a peculiar natureza tecnológica dos engenhos que visavam reproduzir a voz e a fala, o que, do ponto de vista teórico, é assaz relevante, tem que ver com a concepção funcionalista da voz como 'instrumento'. É nos estudos fisiológicos sobre o sistema fonador que encontramos um 'segundo momento' alusivo à íntima convergência da tecnologia com a ciência da voz. Podem servir-nos, a título exemplificativo, as formulações de Charles Bell, o qual, já nas primeira décadas do século dezanove, apela a um estudo fisiológico do fenómeno da voz segundo um amplo critério de integração somática, por meio do qual os órgãos vocais são intimamente relacionados. São as múltiplas relações entre as partes constituintes do corpo humano que, segundo Bell, verdadeiramente o distinguem das "[…] coisas artificiais […]"; evidenciando, nesse aspecto, os órgãos vocais a melhor expressão da "[…] cooperação funcional […]" que anima o corpo humano (Bell, 1832, p. 299).

Embora Bell considere o papel relevante da expressão na formação da inteligência humana – chegando mesmo a afirmar que a expressão "[…] precede as operações intelectuais […]" e a mente humana e as actividades mentais conscientes são condicionadas pelos movimentos espontâneos dos músculos (Bell, 1865, p. 198-199) –, esse papel é, por ele, formulado em pleno acordo com um pressuposto instrumentalista das funções do órgão. Na concepção do autor, a combinação de diferentes funções fisiológicas está na origem do fenómeno da expressão, nomeadamente, é graças à associação do órgão da respiração com o coração que se gera "[…] o instrumento da expressão […]" e a própria visibilidade das emoções por este desenvolvidas (Bell, 1865, p. 86). Logo, apesar dos esforços teóricos de Bell e porque é, em rigor, uma ciência com uma extensa pregnância tecnológica, a fisiologia nunca consegue ultrapassar os seus pressupostos mecanicistas e funcionalistas, os quais lhe permitem conceber os órgãos e o corpo em plena analogia com as estruturas e o uso que fazemos dos artefactos tecnológicos.

Como 'terceiro e último momento', impera referir o legado teórico da linguística moderna, que assimilou, pela mão de Ferdinand de Saussure, o paradigma funcionalista resultante da estreita convergência entre as investigações tecnológicas e os estudos fisiológicos. Mas, se Bell, focado na fundamentação científica da sua teoria da expressão, aborda as dimensões fisiológicas dos órgãos vocais em pleno acordo com as suas possibilidades afectivas, já Saussure rompe com o legado das dimensões expressivas da voz e, considerando, apenas, as dimensões fisiológicas da fonação, depressa é levado a estabelecer uma distinção excludente entre as interacções discursivas e os sistemas linguísticos. Ao não incluírem, positivamente, as qualidades sensíveis introduzidas e mediadas pela percepção nos actos de comunicação e na formação da linguagem, os pressupostos teóricos da linguística saussureana são, na verdade, idênticos aos do dualismo cartesiano 'alma-corpo', o qual se deixa actualizar e traduzir pela dicotomia *langue-parole*. Concebendo a linguagem segundo critérios predominantemente informacionistas, Saussure vê-se confrontado com a necessidade de estabelecer uma série de nexos analógicos entre os sistemas linguísticos e o domínio tecnológico, para contrastar a autonomia da *langue* ante a heteronomia da *parole*. À arbitrariedade do falar é associada a própria natureza

[4] "[…] Denn, hat man eine Sprache durch Handzeichen für das Aug erfinden können, so läßt sich kein Grund dafür finden, warum man nicht auch eine Sprache durch Töne für das Ohr hätte erfinden".

dos órgãos vocais, os quais, segundo o autor, "[...] são tão exteriores à língua quanto os aparelhos eléctricos usados para transcrever o alfabeto Morse são estranhos a esse alfabeto"[5] (Saussure, 2005, p. 36). De facto, o cartesianismo, na sua versão linguística, faz com que a ideia saussureana de *langue* se imponha acriticamente e nos leve a assumir que, quer na fala quer na escrita, os meios materiais e sensíveis da linguagem sejam simples suportes, fisiológicos e tecnológicos, de translação de um sistema autónomo que lhes é estranho.

Os três momentos reflexivos já mencionados permitem-nos, igualmente, compreender muitas das razões que fundamentam a subjugação da análise da linguagem aos conteúdos puramente informativos. Os pressupostos da teoria da linguagem tenderam, quase sempre, a incidir sobre as relações de representação dadas no âmbito semiótico da formação do signo. Nem mesmo a chamada *speech act theory* – a qual pretende contextualizar o sentido das asserções discursivas em diferentes actos performativos – se mostra capaz de ultrapassar a ideia de representação e a consequente redução da linguagem às funções de enunciação proposicionais. A título de exemplo, para os teóricos da análise proposicional da linguagem, o valor da asserção 'As árvores são animais que se alimentam de frutos' é desprovido de correspondência veritativa, porque é falso que as árvores sejam animais. Logo, a mensagem, abstraída dos nexos intersubjectivos que o falar e o ouvir promovem e inscrevem, é, de facto, desguarnecida de uma 'voz'. O 'valor veritativo' das asserções nunca pode expressar o seu 'valor vocal', já que é por meio da inferência do primeiro que se apuram os vínculos da informação com os modos de comunicação. Esta redução discursiva, permitindo a abstracção normativa dos actos de fala, impede-nos, assim, de fazer outra interpretação do exemplo dado anteriormente. Faltar-nos-á, sempre, a percepção da voz de quem pronuncia a frase e de quem a ouve. Ambos os interlocutores podem, em rigor, ter a mútua compreensão de que as árvores não são animais – expressando a asserção um valor veritativo –, mas já não inferida isoladamente do conteúdo proposicional da asserção. O valor vocal da asserção seria a dimensão que faltaria para determinar quer o seu verdadeiro propósito comunicativo quer a existência ou não de propósito – tal tarefa implicaria, antes de tudo, resgatar a percepção do foco na informação verbalizada.

Ora, num contexto de comunicação presencial, o que não é dito pode ganhar expressão no corpo dos interlocutores, nos seus gestos, nas suas entoações, nos seus silêncios, desencadeando novas possibilidades de comunicação e, por extensão, mais comunicação. "Há uma aderência da fala ao silêncio do outro [...]"[6], tal como assevera Ihde (2007, p. 111), uma vez que os intervalos no discurso são preenchidos pela 'fala' do rosto. Aquilo que melhor pode distinguir a comunicação da mera transmissão de informações reside, precisamente, nesta diferença: 'o dito nunca equivale totalmente ao percepcionado', 'nem o percepcionado pode ser reduzido ao dito'. Mas, longe de serem dois eventos indiferentes e desacoplados, ambos contribuem para a formação do contexto de comunicação, entrelaçando a presença dos interlocutores com o espaço físico onde interagem e inscrevendo, na linguagem, os efeitos empíricos da própria interacção.

Consequentemente, uns dos maiores desafios que se impõem à criação e programação dos dispositivos de voz reside na elaboração de simetrias entre o conteúdo comunicativo e o envolvimento perceptivo. Tal não é um fenómeno novo. Com o desenvolvimento da telecomunicação, despontou, sempre, a necessidade tecnológica de reduzir os efeitos da distância por meio da evocação dos efeitos da proximidade. Numa relevante reflexão sobre os processos discursivos nos contextos de conversação, Emanuel A. Schegloff afirma que, apesar das diferenças introduzidas pela tecnologia, os meios de telecomunicação conservam, ainda, aspectos da interacção social. Nas suas palavras, torna-se "[...] possível, hoje em dia, ouvir o telefone para o qual se liga a ser atendido por uma voz humana, mas, apesar disso, sem falar com um ser humano [...]"[7]; acrescentando, em seguida: "Por menor que seja o grau de consolo, podemos observar que, mesmo as máquinas como o atendedor automático, são construídas com base em princípios sociais e não, somente, mecânicos"[8] (Schegloff, 1968, p. 1090). Todavia, a inscrição de traços sociais na programação dos dispositivos é, neste caso, alavancada pela sequenciação da voz em dois registos autonomizados pela tecnologia – o da pergunta e o da resposta. Como não há nenhuma possibilidade dialógica, o atendedor não responde, verdadeiramente, a uma pergunta; apenas se limita a evocar uma função isolada da voz humana, introduzindo, para o efeito, a sugestividade paradoxal de quem responde não pode responder. Bem diferente é, num contexto presencial, expressar a minha indisponibilidade para comunicar, mesmo estando, para a percepção dos outros que me interpelam e observam, fisicamente presente. Há, nestes casos do direito ao silêncio, uma separação provocada entre comunicação e percepção, embora seja balizada pela consciência e pelos efeitos partilhados da presença.

[5] "[…] sont aussi extérieurs à la langue que les appareils électriques qui servent à transcrire l'alphabet Morse sont étrangers à cet alphabet".
[6] "There is an adherence of speech to the silence of the other […]".
[7] "[…] It is possible, nowadays, to hear the phone you are calling picked up and hear a human voice answer, but nevertheless not be talking to a human […]".
[8] "However small its measure of consolation, we may note that even machines such as the automatic answering device are constructed on social, and not only mechanical, principles".

Informação, incorporação, órgão

A maior parte dos estudos teóricos sobre as relações entre os novos meios tecnológicos e o corpo tem sido centrada na questão da representação visual (Balsamo, 1996), dando, nesse sentido, continuidade ao legado analítico suportado quer pela iconografia artística quer pela exposição do corpo nos *mass media*. Mas, tal como na análise da linguagem, o fenómeno da representação mostra-se, *per se*, insuficiente – e, por vezes, equívoco – para agregar todas as dimensões estruturantes dos liames da tecnologia com o corpo humano. Uma das razões fundamentais que nos revelam essas insuficiências tem que ver, sobretudo, com a própria natureza tecnológica dos novos dispositivos de mediação, os quais, já não se encontrando confinados à mera transmissão de informações, requerem e possibilitam uma manipulação activa por parte do utilizador-observador. É, aqui, na esfera da utilização, que se adensam as grandes diferenças entre os meios digitais e os pré-digitais, nomeadamente por causa da reformulação do papel do corpo na formação dos processos de mediação.

Para fazer jus a todas estas transformações e transpor a circularidade semiósica da representação, temos, pois, de retroceder às relações de incorporação entre corpo e tecnologia. A relevância da voz justifica-se, também, pelas possibilidades teóricas que ela acrescenta ao fenómeno da incorporação. É, partindo deste, que podemos traçar uma distinção teórica fundamental entre 'a voz que ouve' – a voz que incorpora o falar e o ouvir do outro – e 'a voz que é ouvida' – a voz que não incorpora o ouvir e o falar do outro.

Na primeira inscrição vocal – a voz que ouve –, as relações incorporadas têm e transportam o corpo como referência, não existindo uma discriminação funcional excludente de quem fala e de quem ouve. É por via dessa dupla condição – de falar e ouvir – que ela gera presença, no sentido em que reconhece e reforça a irredutível existência do outro. Como bem refere Walter Ong (2002, p. 66), a dimensão activa do corpo nas interacções discursivas "[...] não é acidental ou artificial [...]", já que há, na própria constituição das palavras faladas, "[...] modificações de uma situação existencial total, que envolve, sempre, o corpo"[9]. Se a entoação é a pregnância imediata dessa incorporação que excede o contexto verbal e abarca o contexto existencial, então, torna-se "[...] impossível verbalizar, oralmente, uma palavra sem nenhuma entoação"[10] (Ong, 2002, p. 99).

Na sua análise fenomenológica da voz, Ihde introduz o conceito de 'aura auditiva' (*auditory aura*) para referir as peculiaridades sónicas dos processos de incorporação atinentes à comunicação presencial. Na interacção discursiva entre dois interlocutores, os corpos ultrapassam a condição estática de meras extensões físicas, interpenetram-se, invadem-se mutuamente e, graças à verbalização *in loco*, o espaço empírico entre ambos é 'preenchido' por uma presença partilhada. Ao contrário da experiência musical, por meio da qual se verifica uma propensão para a desincorporação, para uma espécie de evasão dos limites traçados pela instrumentação, a presença do outro é intensificada pelo som da sua e da minha voz, gerando-se, com isso, uma forma de incorporação recíproca (Don Ihde, 2007). Estas formulações de Ihde deixam-se melhor compreender, se se acoplar à incorporação um factor projectivo. Dito de outro modo, 'a voz que ouve' incorpora o outro não só por meio do som. O reconhecimento sonoro do outro resulta, simultaneamente, de uma projecção do que é ouvido na própria fonte de emissão – é a voz que brota daquele corpo. Por via dessa projecção, outras dimensões sensoriais são envolvidas e evocadas (Braga, 2019) – ao ouvir é acrescentado o olhar e à palavra é acrescentado o gesto –, formando-se uma imagem integrada do que é percepcionado com as dimensões somáticas do corpo em questão.

É, por isso, que os fenómenos vocais são, exemplarmente, irredutíveis à esfera da comunicação. Por via deles redunda uma expressão cabal do envolvimento da percepção em cada acto comunicativo, mormente no que à potenciação da presença individual dos interlocutores diz respeito. O falar transcende o mero intercâmbio de informações entre um emissor e um receptor, como, equivocamente, tende a ser advogado pelas teorias linguísticas de inspiração saussureana. Pelo corpo que acusa a sua ressonância, a voz do falar é projectada e individuada no corpo donde brota. Daí, também, que os referentes da fala não sejam, apenas, de ordem proposicional e representativa; eles expressam, ao mesmo tempo, os vínculos somáticos da linguagem e a reentrada do corpo dos interlocutores na formação de sentido do que é e não é falado. Acopladas ao teor comunicativo dos símbolos discursivos há, então, por meio do falar, micropercepções dos corpos dos falantes que podem veicular tanto a estabilidade empírica do contexto discursivo quanto a disponibilidade psíquica dos intervenientes.

Na segunda inscrição vocal – a voz que é meramente ouvida –, quebrada a amplificação somática, se impõe o órgão como centro fisiológico de todas as possibilidades e limites tecnológicos do falar e do ouvir. A voz que

[9] "[…] modifications of a total, existential situation, which always engages the body".
[10] "[…] It is impossible to speak a word orally without any intonation".

encerra a possibilidade de ser apenas ouvida é, nesta distinção, a voz sintetizada e reproduzida tecnologicamente. Prestando-se à sua reprodução e sintetização, a voz integra as expectativas dos que a poderão ouvir, antecipando-se ao próprio acto de escuta e modelando, assim, sem a incorporação do falar e ouvir, a própria forma de audição. Dada a inevitável inexistência de alteridade e estando confinada às formulações matemáticas da teoria da informação e à consequente redução da comunicação à tríade 'emissor-mensagem-receptor', a voz desincorporada assume-se, com efeito, como um órgão que vocaliza para um órgão que ouve.

Logo, a voz que é simplesmente ouvida, não me ouve – quer a falar quer a ouvir o outro –, apenas me devolve o que pretendo e posso ouvir. Todos os signos linguísticos que por mim são emitidos, são, por sua vez, convertidos pelos programas da máquina em sinais de comando. O que implica dizer que, funcionalmente, tal conversão pode ou poderia ser feita mediante o uso de sons não-linguísticos reconhecidos pelos programas da máquina. Portanto, o utilizador dos assistentes digitais não 'fala' com a máquina; ele serve-se, sim, da linguagem para activar as funcionalidades dela, quase com o mesmo sentido de, em certos casos do dia-a-dia, se vociferar uma qualquer palavra para despertar a atenção de alguém que nos não ouve nem vê. Como sugerem estudos recentes, as crianças tendem a transmitir "[...] menos informações aos assistentes de voz do que aos humanos"[11] (Aeschlimann, Bleiker, Wechner, & Gampe, 2020, p. 7). Apesar de os comportamentos estritamente sociais servirem de fundo referencial para o uso dos assistentes digitais, potenciando, dessa maneira, a natureza e a função da automação para a qual são programados, há evidência empírica de que, até mesmo, as crianças são capazes de estabelecer diferenças comportamentais claras entre humanos e dispositivos tecnológicos.

A voz dos assistentes digitais – particularmente, aqueles dispositivos que são activados pela voz dos seus utilizadores – é, pois, desprovida da capacidade de incorporar a voz do outro. O que, acima de tudo, ela revela é o vasto fenómeno de 'importação', por parte da máquina, de certos actos humanos – motores e sensoriais –, para poder suportar a constituição das interfaces tecnológicas (Braga, 2020). Um desses actos importados reside, precisamente, no uso da voz dos utilizadores dos artefactos tecnológicos com a função de comando verbal, como meio de activar e controlar as operações por eles mediadas. Não equivalendo a uma incorporação, a importação implica, somente, a tradução programada de um acto humano num determinado sinal de comando dado à máquina. Convém, contudo, salientar que, apesar de a máquina não incorporar as relações que configuram a sua interface, estas são incorporadas pelos seus utilizadores. A circularidade relacional da alteridade é quebrada, no preciso momento em que os utilizadores incorporam os seus próprios actos sem a reciprocidade dos que são apenas importados e traduzidos pela máquina.

Logo, o que significa este desencontro somático entre incorporação e importação para a formação da tecnologia dos assistentes de voz digitais?

Em geral, poder-se-á asseverar que as mediações suportadas digitalmente obedecem, simultaneamente, a 'critérios estéticos de compensação' dos hiatos psicomotores abertos pela natureza tecnológica dos dispositivos. As formas que essa compensação assume são várias, dependendo, em muito, das principais dimensões motoras e sensoriais envolvidas na constituição da interface. No caso dos assistentes de voz, o registo sonoro é programado para criar nexos sugestivos com a memória somática dos utilizadores, nomeadamente mediante a acentuação de certas qualidades vocais que, nas nossas interacções quotidianas, permanecem discretas ou, até mesmo, imperceptíveis. Daí que, na criação dos assistentes tecnológicos, a questão do realismo seja associada à da confiança. O preceito de que a oralidade gera proximidade e a textualidade, distância, é tido em conta para justificar a aparente 'humanização' destes dispositivos e aumentar a sua eficácia: um incremento significativo nas semelhanças antropomórficas dos dispositivos traduz-se, segundo alguns autores (Burgoon et al., 2000), tanto numa maior disponibilidade psíquica para as informações transmitidas quanto numa maior valoração positiva dos conteúdos delas. A linha que une o realismo e a empatia é, porém, ténue. Ainda que, inicialmente, aplicada à robótica, a famosa teoria do *uncanny valley*, de Masahiro Mori, despertou a atenção para o grau de afinidade que os seres humanos manifestam quando são confrontados com dispositivos robotizados com traços antropomórficos. Se o grau de semelhança revelar, simultaneamente, lacunas entre o artificial e o natural – a ponto de haver um efeito de desilusão provocado pelas imperfeições detectadas na mimetização da aparência humana –, então, como defende Mori, o grau de afinidade será potencialmente menor e depressa redundará num efeito de repulsa. Mais facilmente se aceita um dispositivo desprovido de traços antropomórficos do que outro que, assimilando um perfil idêntico ao dos humanos, não atinja um limiar estético de perfeição (Mori, MacDorman, & Kageki, 2012).

[11] "[…] children share less information with voice assistants than they do with humans".

Segundo a análise de Jutta Weber (2005), os estereótipos tecnológicos – nomeadamente os de género – contribuem para a padronização das relações dos utilizadores com os dispositivos automatizados, na precisa medida em que lhes incutem um certo grau de familiaridade decorrente de interacções sociais já modeladas por valores e práticas comuns. A complexidade da tecnologia é como que mitigada pela evocação performativa das relações sociais estereotipadas, gerando-se, no próprio acto de utilização dos dispositivos, um efeito de espontaneidade. A mediação tecnológica do corpo humano torna-se, ainda, mais complexa, quando se trata de criar uma convergência expressiva entre o meio utilizado e o corpo envolvido, capaz de desencadear a fusão sugestiva de ambos e, por consequência, anular as causas e os efeitos artificiais da própria mediação. O rosto 'fotogénico' e a voz 'radiofónica' são dois exemplos singulares da convergência expressiva alusiva ao universo dos *media*. Mas, independentemente das modalidades estéticas que essas formas de fusão possam assumir, o que de mais importante daí convém reter, tem que ver com a necessidade de as funções e os usos dos meios tecnológicos serem suplementados por um excesso de perfeição humana. A voz 'perfeita' e o rosto 'perfeito' só, verdadeiramente, o são, porque se dão às mediações operadas pelos artefactos. A sublimação do órgão redunda, assim, do facto inevitável de se tratar de um órgão que cumpre uma determinada função tecnológica.

Órgão, reprodução, género

A predominância, nas estruturas dos dispositivos e plataformas digitais, de códigos alfanuméricos binários condiciona a natureza e o nível semântico das representações sociais que por ambos são importadas, bem como as formas da sua reinscrição tecnológica. Embora uma interacção social, mediada pela linguagem, não se deixe reduzir nem substituir por um simples gesto de aprovação ou desaprovação, um categórico 'sim' ou um 'não', nos domínios da comunicação, mediados digitalmente, é o valor informativo que, sobremaneira, dita a articulação das possibilidades semânticas dos conteúdos com as possibilidades sintácticas dos meios. A ideia de 'informação' tem, como principal e implícito correlato teórico, a ideia de 'reprodução'. O que é reprodutível só pode ser concebido sob uma moldura funcional, por meio da qual a informação adquire um valor *per se*. A moldura funcional é traçada pela automação do órgão – do falar e do ouvir, no caso dos assistentes de voz digitais – perante as relações somáticas que o envolvem. Porém, estas jamais estão localizadas apenas nos corpos de um 'tu' e um 'nós', como, também, no meu corpo. Como disso nos dá conta Ihde (2007, p. 136), no acto de falar, "[…] eu sinto a minha voz ressoando pelo menos na parte superior do meu corpo"[12]; este fenómeno da 'auto-ressonância' (*self-resonance*) já não ocorre com a gravação da voz, cuja natureza aparenta ser desprovida do "[…] efeito da minha voz na minha estrutura esquelética e muscular"[13].

O processo de assimilação da voz humana pelos assistentes digitais não é, contudo, unidimensional. Na constituição da própria interface (vocal) entre máquina e utilizador, este último também tende, com o hábito, a assimilar as características sintácticas inerentes à programação algorítmica da voz. Tal como, em geral, sucede na utilização das demais tecnologias digitais, a eficácia pretendida obriga a que os utilizadores apreendam as melhoras formas de articular o vocalizar com os comandos que são dados e acolhidos pela máquina, para que ela possa operar em plena concordância com as finalidades desejadas (Natale & Cooke, 2021).

Dos novos dispositivos digitais – nomeadamente aqueles que recorrem a sistemas algorítmicos de modulação e codificação dos comportamentos dos utilizadores – redunda o ideal tecnológico de que a eficácia da máquina depende, simultaneamente, da adaptabilidade e reajustamento das suas configurações no decurso da utilização. Alan Turing, no seu célebre artigo *Computing Machinery and Intelligence* (Turing, 1950, p. 456), introduziu a expressão "[…] *child-machine* […]" ('máquina-criança') para se referir a um processo de simulação tecnológico da mente humana, ancorado num programa computacional aberto ao que ele designa de "[…] educação apropriada". Tal como se ensina uma criança a adquirir determinados comportamentos e competências cognitivas, também acreditava Turing que o mesmo poderia ser feito no âmbito da autoprogamação das máquinas digitais que servem a pretensa simulação da inteligência humana. Este facto pioneiro na história da chamada 'inteligência artificial' revela-nos, porém, relevantes dimensões afectas à formação dos discursos da própria tecnologia digital, a saber – os nexos analógicos entre corpo e máquina. Aliás, na sua proposta do método de aprendizagem, Turing faz uso do princípio de "[…] punições e recompensas […]", associado ao ensino escolar, como uma forma de condicionar a aquisição de informações por parte da máquina (Turing, 1950, p. 457). Embora a máquina 'não sinta' que é punida nem que é recompensada, o princípio tem um valor funcional operativo para os programadores que conduzem o seu processo de aprendizagem.

[12] "[...] I feel my voice resonate throughout at least the upper part of my body."
[13] "[…] the effect of my voice on my skeletal and muscular framework".

Ora, a construção de analogias entre corpo e máquina vai para lá da mera sugestividade antropomórfica e, ao contrário dos pressupostos funcionais de Turing, abrange, igualmente, um envolvimento activo da esfera das emoções. Tal facto é deveras evidente nos usos tecnológicos da voz feminina. Com a introdução da voz feminina nos assistentes virtuais, há um liame subtil que se deixa pensar entre a reprodução operada pela máquina e a reprodução inculcada ao sexo feminino – o dispositivo transforma-se na 'máquina-mãe'. Desse novo estatuto, no qual o artificial sugere um natural estereotipado, redundam nexos de projecção-identificação, por parte dos utilizadores, que têm o órgão – o sexo feminino e o sexo masculino – como ponto de referência primário. Aliadas às expectativas do género, há, também, evocações estereotipadas da faixa etária, da classe social e da pertença territorial da voz que é reproduzida pelos assistentes digitais (Nass, Moon, & Green, 1997). Mesmo que, alegando uma suposta neutralidade de género dos meios tecnológicos, a dicotomia 'feminino-masculino' não seja inteiramente assumida pelos utilizadores dos dispositivos, há, como demonstram alguns estudos empíricos (Nass & Yen, 2012), uma tendência geral de atribuir à voz feminina possibilidades afectivas – mormente na transmissão de conteúdos referentes à vida sentimental dos seres – que não se encontram na voz masculina, a qual é, por vezes, associada à eficaz transmissão de conteúdos técnicos.

Por conseguinte, a 'máquina-mãe' reproduz o órgão sem corpo, a voz sem corpo, o sexo sem corpo – isto é, reproduz o órgão do estereótipo. Nada nos garante, portanto, que, da natureza acústica da voz, reproduzida e automatizada, possa ser inferido um género correspondente, até mesmo segundo a redução binária 'masculino-feminino'. Reproduzido tecnologicamente, qualquer som pode ser manipulado e adquirir características acústicas que, originariamente, lhe não pertencem. A ideia, assaz disseminada, de que a dicção feminina favorece uma melhor percepção discursiva, porque, como defendem alguns autores (Liu & Holt, 2015), há, na articulação das vogais, uma duração relevante que as torna foneticamente mais salientes, transporta já um determinismo sexual que não é, de todo, defensável. Desde logo, porque, independentemente do género, na reprodução digital, a voz é modelada e programada em concordância com a natureza tecnológica do dispositivo, impondo esta à própria dicção, por exemplo, processos de articulação fonéticos com um elevado grau de transparência.

Logo, a incorporação não é, aqui, um fenómeno desenraizado e indiferente aos múltiplos contextos performativos onde ocorre a utilização dos dispositivos. Pelo contrário, na voz que é ouvida e incorporada se inscrevem as funções específicas desempenhadas pelos dispositivos e, por via disso, a projecção-identificação dessas funções nas funções do corpo da própria voz. Cumpre, pois, ainda, implicar as divisões das funções laborais efectuadas segundo critérios de género e sexo. No imaginário social do trabalho – e apesar de algumas, mas ainda escassas, mudanças significativas – a noção de 'assistente' tende a ser vislumbrada com um perfil feminino. Seja por intermédio de vivências presenciais, como as de um aeroporto, ou de vivências remotas, como as mediadas telematicamente, somos, diariamente, confrontados com a voz feminina cumprindo uma função de assistência. 'Ao homem cumpre gerir, à mulher cumpre assistir' – tal poderia ser a formulação sexista que relega o labor das mulheres para a condição de, entre outras, secretárias, assessoras, telefonistas e recepcionistas. Devido a esse imaginário social do trabalho e às suas múltiplas inscrições segundo o género e o sexo, a voz feminina evoca, ainda, um 'corpo de assistência', um corpo que, desde o século dezanove, transporta tanto os cuidados prestados pelas enfermeiras de guerra quanto as informações transmitidas pelas primeiras telefonistas. Já no final do século dezanove, as palavras de um comentador (Hubert, 1889, p. 260) sobre as vantagens económicas do fonógrafo de Thomas Edison são bem ilustrativas desse estatuto laboral das mulheres, nomeadamente quando afirma que "[...] a jovem mulher da máquina de escrever pode imprimir no papel o que seu empregador ditou para o fonógrafo"[14].

Não são, por conseguinte, os critérios acústicos formais, ancorados no imaginário estético da sociedade, os verdadeiros factores determinantes para a predominância da voz feminina nos serviços de apoio aos utilizadores tecnodigitais. A máquina esconde o corpo individuado da voz, mas a voz, na sua função instrumental, evoca já um corpo talhado pela divisão sexual do trabalho e as suas sucessivas manifestações culturais até aos nossos dias. São, no entanto, várias as tentativas de omitir e encobrir os processos históricos da divisão do trabalho. Para alguns teóricos dos novos dispositivos tecnológicos, a 'inteligência artificial' dos assistentes digitais suportados pela voz feminina, "[...] parecerá muito mais real se ela, também, aderir aos padrões de fala femininos"[15] (Hannon, 2016, p. 35). Além da questão mimética, surge, aqui, a concepção de que o uso de certos vocábulos – como, por exemplo, o pronome pessoal 'eu' – se encontra associado a um

[14] "[...] the typewriter girl can print out upon paper what her employer has dictated to the phonograph".

[15] "[...] will seem that much more real if she also adheres to female speech patterns".

estatuto social 'inferior', quer em termos relacionais e cognitivos quer no que à dimensão laboral diz respeito. A solução para estes pretensos desnivelamentos culturais é, então, concebida segundo critérios de uniformização e refinamento da linguagem ou, nas palavras de Charles Hannon (2016, p. 35), "[...] padrões que subvertam ou contornem aqueles que, geralmente, encontramos no mundo [...]"[16] e que, dessa forma, "[...] possam pavimentar uma parte do caminho em direcção a uma sociedade com maior igualdade de género"[17]. Em suma, neste caso, a máquina seria como que maquilhada por uma suposta linguagem inclusiva, apesar de a função instrumental da voz continuar inalterada e associada ao género feminino.

Considerações finais

Para que o seu uso seja eficaz e massificado, a redução de complexidade dos artefactos tecnológicos é um dos princípios mais relevantes que se lhes impõem. Obedecendo a múltiplos critérios materiais, operativos e estéticos, tal princípio pressupõe, desde logo, um estabelecer de relações entre cada novo artefacto e os precedentes. Se essas relações exibem um perfil simétrico é porque há características dos precedentes que, com maior ou menor grau de transformação, migram para os artefactos subsequentes. Da mesma maneira que o teclado com caracteres da máquina dactilográfica serviu de modelo da interface 'utilizador-computador', também o estereótipo da voz feminina, alimentado pelas primeiras tecnologias de voz, transitou para os assistentes digitais. Para este caso, vale, então, a formulação seguinte: a redução de complexidade e o incremento de eficácia tecnológicas incluem a reprodução e exponenciação do estereótipo.

A habitual feminização dos assistentes digitais vai, por via disso, para lá das qualidades estéticas associadas à voz feminina. Ou melhor, por intermédio da voz, são desencadeados nexos referenciais, conscientes e inconscientes, entre assistência e subserviência, os quais reproduzem, de forma reiterada, muitos dos estereótipos inculcados ao perfil e ao papel das mulheres na sociedade. A crescente expansão da tecnologia digital e a multiplicação de serviços por ela suportados, coroados pelos interesses económicos, são factores decisivos no modo como os estereótipos são actualizados e reintegrados na vida e nas interacções sociais. O que ocorre na esfera – pública e privada – da assistência digital está longe de tipificar um evento isolado e autónomo em face de todos os que ocorrem nas outras esferas de mediação. Como há muito sucede no domínio laboral dos *media* audiovisuais – nos quais a extrema exposição do corpo feminino tende a configurar a silhueta visível da informação –, também, nos assistentes digitais, é à voz feminina que cumpre dar a informação, enquanto quem gera e gere a informação exibe o perfil masculino de 'programador'. Na verdade, o combate à desigualdade de género continuará a ser improfícuo e, deveras, equívoco, se for norteado por princípios unicamente quantitativos de ocupação do espaço de acção. Inversamente, tanto a estratificação quanto a estereotipação laborais exigem respostas qualitativas, capazes de fomentar o equilíbrio na distribuição e no desempenho das funções exercidas. De pouco ou nada servirá, haver maior participação feminina nas forças de produção, se essas forem, profissional e socialmente, consideradas inferiores.

As relações de poder, alicerçadas na divisão sexual do trabalho, não são, por conseguinte, fenómenos suplantados pela sociedade contemporânea; elas persistem, exibindo, sempre, novas formas e novos domínios de aplicação. Ao invés de assinalar uma mudança de rumo, o domínio da assistência digital tem reiterado, até aos dias de hoje, as várias versões laborais decorrentes e nutridas pelo estereótipo da 'mãe-cuidadora', com o pendor de se tratar de um domínio em que, apesar de os rostos da mão-de-obra permanecerem invisíveis, a voz se torna facilmente evocativa da condição feminina. A estabilização dos estereótipos visuais pelos estereótipos auditivos vem, neste caso, adensar o princípio segundo o qual 'a robustez social dos estereótipos em muito depende dos efeitos de convergência entre as modalidades sensoriais que os vinculam à nossa percepção'. As tecnologias digitais têm, nas sua bases materiais e estruturais, esse poder de convergência, o que as torna potencialmente mais permeáveis à replicação do que ainda subsiste como valor inquestionável.

Referências

Aeschlimann, S., Bleiker, M., Wechner, M., & Gampe, A. (2020). Communicative and social consequences of interactions with voice assistants. *Computers in Human Behavior, 112*, 106466. DOI: https://doi.org/10.1016/j.chb.2020.106466

[16] "[…] patterns that subvert or circumvent those we find more generally in the world [...]".

[17] "[…] might pave on part of the road toward a more gender-equality society".

Balsamo, A. M. (1996). *Technologies of the Gendered body: reading cyborg women*. London, GB: Duke University Press.

Bell, C. (1832). Of the organs of the human voice. *Philosophical Transactions of the Royal Society of London, 122*, 299-320.

Bell, C. (1865). *The anatomy and philosophy of expression: as connected with the fine arts* (5th ed.). London, GB: Bohn.

Braga, J. (2019). Imagination, multimodality, and sound. In M. Grimshaw-Aagaard, M. Walther-Hansen, & M. Knakkergaard (Eds.)*, The oxford handbook of sound and imagination, Volume 1* (p. 131-148). New York, NY: Oxford University Press.

Braga, J. (2020). *Teoria das formas imagéticas. Ensaios sobre arte, estética, tecnologia*. Coimbra, PT: Grácio Editor.

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Cederberg, C., Lundeberg, M., & Allspach, L. (2000). Interactivity in human–computer interaction: a study of credibility, understanding, and influence. *Computers in Human Behavior, 16*(6), 553-574. DOI: https://doi.org/10.1016/S0747-5632(00)00029-7

Dudley, H. (1940). The carrier nature of speech. *The Bell System Technical Journal, 19*(4), 495-515. DOI: https://doi.org/10.1002/j.1538-7305.1940.tb00843.x

Hannon, C. (2016). Gender and status in voice user interfaces. *Interactions, 23*(3), 34-37. DOI: https://doi.org/10.1145/2897939

Hubert, P. G. (1889). The new talking-machines. *The Atlantic Monthly, 63*(376), 256-261.

Ihde, D. (2007). *Listening and voice: phenomenologies of sound* (2nd ed.). New York, NY: University of New York Press.

Liu, R., & Holt, L. L. (2015). Dimension-based statistical learning of vowels. *Journal of Experimental Psychology: Human Perception and Performance, 41*(6), 1783-1798. DOI: https://doi.org/10.1037/xhp0000092

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics & Automation Magazine, 19*(2), 98-100. DOI: https://doi.org/10.1109/MRA.2012.2192811

Nass, C., & Yen, C. (2012). *The man who lied to his laptop: what our machines can teach us about human relationships*. New York, NY: Current.

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology, 27*(10), 864-876. DOI: https://doi.org/10.1111/j.1559-1816.1997.tb00275.x

Natale, S., & Cooke, H. (2021). Browsing with Alexa: interrogating the impact of voice assistants as web interfaces. *Media, Culture & Society, 43*(6), 1000-1016. DOI: https://doi.org/10.1177/0163443720983295

Ong, W. J. (2002). *Orality and literacy: the technologizing of the word*. New York, NY: Routledge.

Saussure, F. (2005). *Cours de linguistique générale*. Paris, FR: Payot.

Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist, 70*(6), 1075-1095.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: The University of Illinois Press.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, LIX*(236), 433-460. DOI: https://doi.org/10.1093/mind/LIX.236.433

von Kempelen, W. (1791). *Mechanismus der menschlichen Sprache nebst der Beschreibung seiner sprechenden Maschine*. Wien, AT: Degen.

Weber, J. (2005). Helpless machines and true loving care givers: a feminist critique of recent trends in human-robot interaction. *Journal of Information, Communication and Ethics in Society, 3*(4)*,* 209-218. DOI: https://doi.org/10.1108/14779960580000274