



## O audiolivro e a inteligência artificial “leitora”: fronteiras intermídiais

*The audiobook and the reading by an artificial intelligence:  
intermedial frontiers*

Jaimeson Machado Garcia <sup>a, \*</sup> 

Ana Cláudia Munari Domingos <sup>a</sup> 

Rejane Frozza <sup>a</sup> 

**RESUMO:** O audiolivro é um tipo de mídia que, historicamente, demanda da gravação da leitura de um texto em voz alta capaz de ser reproduzida para ser midiado. Quando essa “leitura”, no entanto, é feita por uma Inteligência Artificial, a exemplo da assistente virtual Alexa, ainda podemos compreender esse produto de mídia sonoro como um audiolivro? Entendendo essa pergunta como uma questão que emerge de novas relações entre as mídias digitais, o presente artigo tem por objetivo definir que tipo de fenômeno é esse. Para isso, buscamos nos Estudos de Intermidialidade, a partir dos modelos propostos por Lars Elleström (2021), junto, ainda, a outros autores que nos auxiliaram a preencher determinadas lacunas e aprofundar outras de sua teoria, os elementos necessários para essa análise. Para ser possível aprofundarmos tal discussão, tomamos de exemplo o e-book e o audiolivro de *O Alquimista* (COELHO, 2017 [1988]), de Paulo Coelho, para a comparação entre a leitura realizada por um humano e a “leitura” realizada por uma Inteligência Artificial. A partir da análise, concluímos que a “leitura” realizada pela Alexa é, na verdade, uma oralização (BAJARD, 2014), o que corresponde a um processo de decodificação sonora das palavras. Assim, quando comparada a um audiolivro lido por um ser humano, resultado de uma transmediação, a decodificação da Alexa se trata de uma “re-mídiação”, isto é, uma re-exibição das modalidades material, espaçotemporal, sensorial e potencialmente semióticas de um e-book por meio de uma mídia técnica de exposição diferente.

**Palavras-chave:** Audiolivro; E-Book; Inteligência Artificial; Intermidialidade; Re-Mídiação.

**ABSTRACT:** Historically, the audiobook is a type of media that requires recording the reading of a text aloud that can be reproduced in order for it to be mediated. However, what happens when this “reading” is done by Artificial Intelligence in real-time, instantaneously, such as by means of a virtual assistant like Alexa? Can this mediation be seen as an audiobook? With the view that this question is an issue that emerges with the new relations among digital media, the aim of this article is to define what kind of phenomenon this is. Hence, we use Intermidiality Studies based on the models proposed by Lars Elleström (2021) and other authors who help us fill in and develop specific gaps in the theory, to find the needed support for this analysis. To expand our discussion, the example used was the e-book and audiobook of *The Alchemist*, by Paulo Coelho, to compare the reading done by a human being and the “reading” done by Artificial Intelligence. Based on our analysis, the conclusion is that the “reading” carried out by Alexa is in fact an “oralization” (BAJARD, 2014), corresponding to a process of audio decoding of words. Thus, compared to an audiobook read by a human, the result of a transmediation, Alexa’s decoding is a “re-mediation”. In other words, it is a re-exhibition of the material, spatial-temporal, sensorial, and potentially semiotic modalities of an e-book by means of a different technical media of display.

**Keywords:** Audiobook; E-Book; Artificial Intelligence; Intermidiality; Re-Mediation.

<sup>a</sup> Programa de Pós-Graduação em Sistemas e Processos, Universidade de Santa Cruz do Sul, Santa Cruz do Sul, RS, Brasil.

\* Correspondência para/Correspondence to: Jaimeson Machado Garcia. E-mail: jaimesonmachadogarcia@gmail.com.

Recebido em/Received: 14/03/2023; Aprovado em/Approved: 25/05/2023.

Artigo publicado em acesso aberto sob licença [CC BY 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/) 

## AS FRONTEIRAS INICIAIS

A presença cada vez mais constante das Inteligências Artificiais (I. A.) em nosso cotidiano vêm ocasionando o importante debate sobre os limites da interação homem-máquina, que trazem à tona as potencialidades tecnológicas e os limites éticos de algumas delas. O ChatGPT, sistema baseado no *deep learning* (aprendizagem profunda, em português) cujo foco até então está na interação verbal com seus usuários, é uma dessas recentes inteligências que coloca em xeque alguns agenciamentos humanos em relação ao maquínico.

O debate, no entanto, é mais amplo, e faz gerar reflexões sobre a forma como produzimos, distribuímos e consumimos os mais variados produtos de mídia. No campo editorial, várias são as transformações em curso a partir das novas tecnologias e também da I. A.. Como averiguar, por exemplo, a autoria de um texto publicado por meio de plataformas de autopublicação como o Kindle Direct Publishing, da Amazon, ou até mesmo durante a fase contratual com uma editora, tendo disponível uma ferramenta dotada com a funcionalidade de criar textos originais, a exemplo do ChatGPT?

Além do debate ético e mesmo jurídico, um outro aspecto das questões editoriais em torno das I. A., que neste artigo nos interessa particularmente, é a do desenvolvimento da capacidade de transformar o texto verbal escrito para o texto oralizado de maneira cada vez mais eficiente<sup>1</sup>. Compreendida como uma alternativa econômica ou até mesmo como uma possível substituta da vocalização humana na concepção de audiolivros, essa habilidade das I. A. tem-se tornado uma realidade cada vez mais frequente.

Em plataformas de *streaming* de áudio como o Audible, o oferecimento de produtos de mídia cujos agentes da voz<sup>2</sup> são I.A. — identificadas de maneira abrangente como “voz sintética”<sup>3</sup> – tem-se tornado cada vez mais comum. Afora as distribuições por serviços online como esse, inteligências artificiais voltadas à automação doméstica e

---

<sup>1</sup> Preferimos chamar a esse processo de oralização, em vez de vocalização, termo que tem sido às vezes utilizado. No entanto, talvez o termo mais eficiente para distinguir voz humana de voz sintética seria sonorização.

<sup>2</sup> Em outro estudo em desenvolvimento, observamos uma dispersão conceitual em torno do responsável por transmediar um texto escrito para sua versão sonora e, por isso, utilizamos a expressão agente da voz, embora o termo previsto por nós que talvez se encaixa mais efetivamente nesse contexto seja “ledor”, pelo fato de não ler com a intenção de alcançar o sentido do texto (o que é feito anteriormente, em silêncio, no processo de apreender o texto), e sim de decodificá-lo na “re-midiação” da versão escrita para uma versão sonora. Já o termo “re-midiação” será elaborado ao final do trabalho, mostrando a diferença entre nossa perspectiva e o termo “remediação”, de Bolter e Grusin (2000).

<sup>3</sup> Disponível em:

[https://www.audible.com/search?searchNarrator=Voz+Sint%C3%A9tica&ref=a\\_search\\_c3\\_INar\\_rator\\_1\\_1\\_1&pf\\_rd\\_p=83218cca-c308-412f-bfcf-90198b687a2f&pf\\_rd\\_r=7YK16FDEYS17B5PA49VN&pageLoadId=MqMUJtnNjqGd1LQe&creativeId=od6f6720-f41c-457e-a42b-8c8dceb62f2c](https://www.audible.com/search?searchNarrator=Voz+Sint%C3%A9tica&ref=a_search_c3_INar_rator_1_1_1&pf_rd_p=83218cca-c308-412f-bfcf-90198b687a2f&pf_rd_r=7YK16FDEYS17B5PA49VN&pageLoadId=MqMUJtnNjqGd1LQe&creativeId=od6f6720-f41c-457e-a42b-8c8dceb62f2c). Acesso em: 7 de mar. de 2023.

ao gerenciamento pessoal, presentes nos chamados alto-falantes inteligentes, também são capazes de realizar tarefas semelhantes ao acessarem bibliotecas virtuais ou ambientes digitais onde se encontram armazenados arquivos de textos dos usuários, como os ebooks.

Dentre elas podemos citar as que são classificadas e comercializadas como assistentes virtuais, presentes em computadores, smartphones ou alto-falantes inteligentes, como é o caso da Siri (Apple), da Cortana (Microsoft), do Google Assistente (Google), e aquela que é cerne do nosso estudo: a Alexa (Amazon). Inserida em uma complexa estrutura que se estende do comércio eletrônico de produtos a uma cadeia de serviços da Amazon, como o próprio Audible, a Alexa tem como principal aplicabilidade, assim como suas concorrentes, a automação residencial e o gerenciamento pessoal por comando de voz, os quais englobam ações como abrir e fechar cortinas, acender e apagar as luzes, lembrar de compromissos, responder a perguntas pesquisáveis na internet, informar notícias e a previsão do tempo, tocar músicas ou até mesmo contar piadas quando solicitada.

Contudo, nosso interesse nessa I.A. em específico se deu, justamente, pela sua habilidade de “ler”<sup>4</sup> em voz alta os ebooks que se encontram armazenados na biblioteca Kindle do usuário. Essa prática de consumo de livros eletrônicos por meio de sistemas falantes como esse fez com que nos indagássemos se essa “leitura” realizada de maneira autônoma e em tempo real é equiparável a um audiolivro. Caso seja, a interpretação levantada por nós é a de que esse fenômeno se caracteriza como um tipo de mídia tão semelhante ao tipo de mídia audiolivro que suas fronteiras são indissociáveis entre si. Mas caso não, como poderíamos, então, defini-lo?

A fim de encontrarmos uma resposta a esse questionamento, buscamos nos Estudos de Intermidialidade subsídios que respaldem a nossa análise, visto que estamos tratando de relações entre mídias. Dentre os autores que por meio dessa perspectiva objetivaram conceituar as maneiras como essas relações se dão, destacamos a teoria proposta por Lars Elleström (2021, p. 11), que parte do pressuposto geral de que “todas as mídias são multimodais e intermediais no sentido em que são compostas por múltiplos recursos básicos e só podem ser completamente compreendidas em relação a outros tipos de mídia com as quais compartilham recursos básicos”. A Intermidialidade, é preciso dizer, além de desconstruir as mídias em suas semelhanças e diferenças, torna-se um discurso sobre as mídias, fornecendo modelos e terminologia para a discussão. Como um campo que atravessa diferentes espaços de conhecimento, o aporte de outras áreas de estudo é natural para a Intermidialidade e, nesse caso que inclui as I.A., não é diferente.

Em síntese, esse trabalho tem por intuito traçar um paralelo entre a “leitura” de uma I.A. e o audiolivro enquanto dois tipos de mídias. E, para ser possível aprofundarmos tal discussão, selecionamos como exemplo um produto de mídia que permitisse essa

---

<sup>4</sup> Ao longo desse estudo, iremos nos referir à “leitura” da Alexa entre aspas até o momento em que definimos que tipo de fenômeno a assistente virtual da Amazon realiza.

comparação, o que acabou culminando na escolha de *O Alquimista* (COELHO, 2017 [1988]; 2021) em suas versões digitais em e-book e audiolivro, tendo este último como agente da voz a atriz e escritora brasileira Beth Goulart.

A escolha por esse título em específico se deu pelo fato de que ambos os tipos de mídias — a “leitura” do e-book pela Alexa e o produto de mídia audiolivro — podem ser percebidos por meio de uma mesma mídia técnica de exposição externa: a dos alto-falantes da família Echo, considerados “inteligentes” por sua integração à Alexa. Enquanto o e-book estava à venda no principal site de comércio da Amazon e é armazenado na biblioteca virtual do Kindle após a compra, o segundo se encontrava disponível no Audible, uma das subsidiárias da empresa com foco no *streaming* de produtos de mídia baseados exclusivamente em ondas sonoras. Tal seleção foi possível pois, embora voltado majoritariamente aos consumidores estadunidenses até a conclusão desse estudo, a Audible possuía um significativo número de opções em língua portuguesa capazes de serem acessadas no Brasil pela Alexa, dos quais estava disponível, justamente, *O Alquimista* (2021).

É importante observar que nossa análise desse produto de mídia se volta aos aspectos de sua midialidade, uma vez que seu conteúdo não nos interessa em sua dimensão literária, ou seja, sua linguagem e conteúdo fabulístico. Outro fator de escolha foi a presença de sons organizados e sons não verbais<sup>5</sup>, que, embora não sejam elementos inerentes a todos os audiolivros, são importantes na compreensão dos fenômenos que aqui nos interessam.

Definidos esses critérios de seleção, o presente estudo se encontra estruturado em três seções: na primeira, apresentamos um panorama geral sobre algumas características das I.A. que podemos encontrar na Alexa; na seção seguinte, discorremos sobre as questões que envolvem os fenômenos de transmidiação, por meio do modelo de Elleström (2021); por fim, na terceira e última seção, verificaremos se essa “leitura” realizada pela Alexa é equiparável a um audiolivro, elencando, posteriormente, nossa análise desse resultado.

## **UMA FRONTEIRA MAIS COMPLEXA: AS INTELIGÊNCIAS ARTIFICIAIS “FALANTES”**

No livro *Performance, recepção, leitura*, publicado pela primeira vez no ano de 1990, o pesquisador e crítico literário Paul Zumthor (2018, p. 16, grifo nosso) escreveu que, no

---

<sup>5</sup> Com base na perspectiva de autores como Signe Kjær Jensen e Niklas Salmose (2022), a partir dos Estudos de Intermidialidade, sons organizados se referem à música, enquanto os sons não verbais, aos efeitos sonoros. Nesse sentido, os sons organizados, quando acompanhados de canções (as palavras cantadas) podem ser entendidos como sons organizados verbais. Vejamos que a inclusão de uma canção pode ser interpretada como elemento diegético, já que é feita da mesma “matéria” da narrativa — as palavras. Deixaremos para aprofundar a relação entre a performance vocal, os sons organizados e os sons não verbais em um estudo subsequente, no qual focalizaremos as fronteiras entre o audiolivro e o cinema, suas aproximações e distanciamentos.

futuro, a “abstração vocal será tanto maior que já não se tratará de gravação, mas de voz fabricada”. Naquela época, entretanto, a reprodução dos processos de geração de fala por meio de modelos articulatórios sintéticos semelhantes ao trato vocal humano já não era apenas uma previsão, mas uma realidade na área da computação, com o desenvolvimento contínuo de máquinas eletrônicas dotadas de *affordances*<sup>6</sup> como essas.

Ao estabelecermos um breve retrospecto sobre dispositivos dotados de um certo tipo de inteligência como são os alto-falantes Echo, encontramos registros que datam desde o século XVIII, quando o desenvolvimento de autômatos havia se tornado popular. Ao retomarem esse percurso histórico, Ben Gold, Nelson Morgan e Dan Ellis (2011) contam que um dos primeiros dessa natureza foi construído pelo inventor húngaro Wolfgang von Kempelen (1734-1804), o mesmo a criar a polêmica máquina de jogar xadrez dotada de uma suposta inteligência artificial chamada de O Turco. Esta, considerada em um primeiro momento como uma das primeiras máquinas pensantes da história, logo acabou se revelando uma farsa após a confirmação da suspeita de que, de alguma forma, havia uma interferência humana vinda de dentro da mesa na qual o autômato ficava apoiado.

Assim como a de xadrez, a máquina falante de Kempelen também requeria a ação manual de uma pessoa. Feita de objetos que já existiam para outros fins, cada um dos componentes presentes em sua estrutura tinha por objetivo reproduzir as funções do conjunto de órgãos responsáveis pela fonação humana, como um fole para a função dos pulmões, tubos de latão para as narinas, canos de junco para a glote, um funil de borracha para a boca e uma caixa de madeira para o tórax. Quando esses componentes eram acionados de maneira sincronizada tal qual a um instrumento musical, a máquina falante de Kempelen era capaz de emitir sons que se assemelhavam aos da fala humana, ainda que de maneira bastante precária devido à impossibilidade de reproduzir determinados fonemas. Mas, por mais que estivesse envolto por determinadas limitações técnicas, máquinas falantes como essa acabaram abrindo caminhos para outras de mesmo tipo com melhores resultados. Dentre elas, os autores destacam a Voice-Operating Demonstrator, a qual ficou popularmente conhecida pela sua sigla, Voder, por ter sido a primeira eletrônica. Desenvolvida pela Bell Telephone Laboratory, a Voder foi apresentada de maneira performática pela primeira vez em 1939, durante a Feira Mundial de Nova York, com o intuito de recepcionar os visitantes ao dar-lhes boas-vindas. Parecido ao formato de um piano, seu funcionamento também dependia da ação de um operador que, para conseguir sincronizar as teclas, botões e pedais que o compunha, precisava ser treinado no decurso de seis meses a um ano devido à complexidade de sua operação. (Gold, Morgan, Ellis, 2011)

---

<sup>6</sup> Em *O conceito de affordance como estratégia generativa no design de produtos orientado para a versatilidade*, José Carlos Broch (2010, p. 25) define *affordance* como: “[...] o produto das relações entre estruturas físicas do ambiente e o intelecto dos seres vivos”. No caso, o ambiente pode ser um objeto.

Apesar de ter se tornado uma das principais referências para outras de mesma finalidade, a interferência humana foi se tornando cada vez menos necessária à medida que a computação foi evoluindo em virtude do desenvolvimento de processadores mais rápidos e armazenamento de dados maiores, culminando no surgimento de sistemas cada vez mais autônomos e simultâneos. Para Xuedong Huang, Alex Acero e Hsiao-Wuen Hon (2001), uma das principais motivações para o contínuo desenvolvimento dessas máquinas capazes de “falar” — e também de “compreender” a fala —, estava na busca por novas interações entre seres humanos e computadores como uma alternativa à Interface Gráfica do Usuário. Ou seja, a busca de melhorias se concentrava em formas de substituir a interação realizada por meio da representação gráfica de desenhos, imagens, sinais e pistas expostas na tela a partir dos cliques do mouse, da digitação do teclado ou, ainda, através do toque dos dedos, gerando um feedback instantâneo das tarefas e processos que estão sendo realizados, conforme explica Lucia Santaella (2013) sobre esse tipo de navegação.

Em dispositivos que funcionam exclusivamente por meio do comando de voz, caso dos alto-falantes inteligentes da família Echo ou da Interface Gráfica do Usuário voltada à acessibilidade de pessoas com deficiência visual, presentes em alguns computadores, Huang, Acero e Hon (2001) explicam que são necessárias outras *affordances*, as quais devem englobar ao menos três subsistemas.

Tais subsistemas são compreendidos entre: 1) o que reconhece a fala humana e a converte em texto; 2) o que realiza o procedimento contrário, ou seja, converte o texto em a fala; e 3) o que “compreende” a língua falada, mapeia as palavras e as transformam em ações. Por mais simples que possam parecer quando assim sintetizados, cada um deles se encontra integrado a uma complexa rede de fatores relacionados à voz humana (como a ritmo, entonação e demais atributos da fala), linguísticos (prosódia e fonética) e computacionais (a programação em si) que, em conjunto, possibilitam o seu funcionamento. Vejamos que, quando falamos em competências tais como “ler” e “compreender”, mantemos essas palavras entre aspas, indicando que talvez não possamos entendê-las como *affordances*, visto que são, até então, habilidades humanas.

Embora esses três subsistemas integrem a Alexa, há somente um que nos interessa neste estudo: o *text-to-speech* — o qual podemos traduzir como texto-para-fala, para o português —, ou simplesmente TTS, como utilizaremos a partir daqui. O TTS faz a conversão de textos escritos, que são processados na sequência fonética correspondente, seguida da análise prosódica que anexa informações de altura e duração. Em concomitância, um outro componente de síntese de fala decodifica esses parâmetros totalmente demarcados e gera uma onda de fala correspondente (Huang, Acero e Hon, 2001).

Segundo Paul Taylor (2009), um dos primeiros usos reais do TTS foi o de oferecer às pessoas com algum tipo de deficiência visual uma alternativa de leitura quando não era

possível ter acesso às versões em braile ou um vidente<sup>7</sup> pudesse fazer a leitura em voz alta. No entanto, sua funcionalidade acabou sendo apropriada por áreas como a do telemarketing, cada vez mais automatizada por questões práticas e comerciais. Para o autor, entre as razões para esse interesse está a possibilidade de reproduzir mensagens que não foram pré-gravadas, oportunizando que as máquinas falem qualquer coisa, “[...] independentemente de a mensagem desejada ter sido originalmente falada ou não” (Taylor, 2009, p. 3, tradução nossa)<sup>8</sup>.

Pensando a partir desse subsistema, em concomitância com as proposições teóricas de Elleström (2021), é necessário considerarmos ao menos três relações fundamentais para a Alexa, que nos trazem os subsídios necessários para analisarmos se a sua “leitura” é uma transmidiação: 1) a da produção da voz sintética e da percepção; 2) a da performance e do corpo; 3) e da consciência e inteligência. A primeira delas e, possivelmente, a mais evidente, é trazida pela conjuntura entre a perspectiva teórica de Rabiner (2010) em concomitância à de Michel Chion (2012).

Na visão do primeiro, há diferentes parâmetros que devem ser considerados para que o sistema TTS seja eficiente, dentre os quais destacamos a inteligibilidade e a naturalidade. O primeiro diz respeito à capacidade do perceptor de conseguir entender a mensagem transmitida pela voz sintética. No caso, se durante a escuta da “leitura” realizada pela Alexa ele consegue decodificar as palavras, ou mais especificamente, se ele consegue perceber a midiação da palavra como signo simbólico. Caso seja, é muito provável que ele seja capaz de compreendê-las com a mesma habilidade de quando um texto é vocalizado por um ser humano.

Essa diferença entre decodificar e compreender tem relação com a distinção entre “midiação” e “representação”, como colocada por Elleström (2021), a qual vale a pena citarmos aqui: enquanto a midiação é um processo que desemboca na percepção do que está sendo comunicado, a representação, que ocorre de maneira quase simultânea, o perceptor passa a dar sentido ao valor cognitivo<sup>9</sup> midiado.

No entendimento de Taylor (2009) sobre tais parâmetros, a inteligibilidade foi um problema até a década de 1970, quando passou a ser possível desenvolver falas inteligíveis pelos sistemas TTS. Na época, Taylor (2009) argumentou que o problema estava justamente na naturalidade que, como o próprio vocábulo indica, consiste em tornar a voz sintética tão natural que passe a ser indistinguível da voz humana. Ou melhor dizendo, que a sintética seja percebida “como se” humana. Mas podemos dizer que, hoje, as adversidades que envolviam a criação de naturalidade nas vozes sintéticas quase foram totalmente superadas.

---

<sup>7</sup> Nos estudos voltados à acessibilidade de pessoas com deficiência visual, a nomenclatura “vidente” é utilizada para denominar aqueles que veem.

<sup>8</sup> Em inglês: “[...] *regardless of whether the desired message was originally spoken or not*”.

<sup>9</sup> Valor cognitivo é aquilo que é transferido entre as mentes durante o ato comunicacional.

Ainda que seja possível mimetizar a fala humana, no entanto, o aparelho fonador é um sistema orgânico extremamente complexo. E, assim como outras inteligências artificiais de uso doméstico, a Alexa possui determinadas limitações em seu subsistema TTS, as quais podemos definir a partir dos diferentes tipos de escuta humana. Tomando como base os estudos de Pierre Schaeffer, Chion (2012) as divide em três categorias, as quais são reorganizadas por Santaella (2005) em virtude das correspondências evidentes com as categorias peirceanas.

Levando em conta objeto, representamen e interpretante, a primeira categoria é a da escuta reduzida. Ela diz respeito ao próprio som em si enquanto objeto, e não enquanto veículo para uma outra coisa. Trata-se, portanto, da escuta de caráter estético, que objetiva entender aquele som *per se*, invocando-lhe características. Nesse sentido, não basta que esse som seja escutado uma única vez, pois precisa ser analisado e, para ser analisado, é preciso que seja gravado, visto que não é possível reproduzi-lo da mesma forma como foi da primeira vez.

Já a segunda categoria é a da escuta causal, considerada como a mais comum entre as três. Concordando com a perspectiva de Chion, David Sonnenschein (2001) explica que a escuta desse tipo ajuda a identificar se o som advém de voz humana, animal não humano, máquina, instrumento musical ou de alguma outra fonte geradora de ondas sonoras. Ou seja, como a própria nomenclatura deixa ver, esse tipo de escuta busca identificar a natureza e as fontes do som escutado.

Por fim, a terceira categoria é a da escuta semântica, que diz respeito ao código ou a linguagem necessária para compreender e interpretar o som. Por essa razão, Santaella (2005, p. 86) argumenta que “[...] trata-se de uma escuta puramente diferencial, pois a língua se estrutura como um sistema de oposições [...]”, e complementa dizendo que: “[...] um fonema não vale por seu valor acústico absoluto, mas pelos traços distintivos que estabelece com outros fonemas”.

Em relação à Alexa, a escuta reduzida possibilita compreendermos que essa limitação se dá, em especial, no que diz respeito à prosódia, como a ausência de inflexões ou outros tipos de inferências na voz, como veremos ao longo da análise. Ainda assim, como podemos adiantar, sua “leitura” linear passa a impressão de que se trata de uma pessoa transmitindo oralmente em tempo real um texto quando observada por meio da escuta causal, por mais que a fonte de origem seja um alto-falante inteligente como o Dot — modelo sem tela da família Echo. Já a escuta semântica somente é possível ser analisada a partir do produto de mídia, pois envolve a interpretação da linguagem verbal como signo.

Já a segunda relação é a do corpo biológico e a da leitura em voz alta, visto a necessidade de traçarmos um paralelo com um audiolivro transmitido sonoramente por um ser humano. Para isso, nos apropriarmos de dois teóricos que julgamos importantes para esse estudo: Zumthor (2018), citado na abertura dessa seção, e o francês Élie Bajard (2014), que segue por um caminho semelhante ao do primeiro.

Partindo da ideia de que a decifração é um ato neutro, visto que seu principal objetivo é a obtenção de informações, Zumthor (2018, p. 35) argumenta que para “[...] que um texto seja reconhecido por poético (literário) ou não depende do sentimento que nosso corpo tem”. Entendendo-a como um acontecimento oral e gestual, o referido teórico denomina a performance como uma realidade experimentada ao reger, simultaneamente, a finalidade de transmissão, o tempo, o lugar, a ação do locutor e, em uma ampla medida, a resposta do público — elementos esses considerados mais importantes do que as regras postas na obra enquanto uma sequência de frases, responsáveis por engendrar o contexto real e determinar o alcance.

Para Zumthor (2018), existem diferentes graus performáticos: a leitura solitária seria algo em torno próximo a zero, enquanto a audição acompanhada da visão seria a mais completa de todas. Mas, quando há falta de um elemento de mediação, como a visão, a oposição entre uma leitura em voz alta e uma performance tende a diminuir, visto que, para ele, a performance não pode ser reduzida ao estatuto de objeto semiótico, pois há algo nela que transborda a ponto de recusar a funcionar como signo. A performance acontece nessa relação de presença entre os corpos, incluindo os gestos, o lugar e o cenário. Ainda que essa performance seja mediada apenas através do som, Zumthor (2005) entende que esses elementos permanecem em algum grau nessa vocalização.

Já Bajard (2014) parte da problemática de que a leitura em voz alta foi a primeira modalidade em exercício desde a origem do alfabeto, seguida tardiamente pela leitura silenciosa. Em vista do longo percurso histórico da escrita, o entendimento sobre a transformação dos signos escritos em signos falados acaba por vezes se entrelaçando e se confundindo ao tratarem de processos contíguos, porém com nomenclaturas diferentes.

Em vista disso, o teórico retoma algumas conceitualizações já fundamentadas e propõe outras novas a partir de seu entendimento sobre o assunto. Para ele, há três etapas que envolvem a leitura, sendo a primeira a decifração — correspondente à decodificação de Zumthor (2018) — cujo objetivo é o de extrair a essência das mídias verbais escritas. Já a segunda etapa é a da leitura corrente, que permite a impregnação de sentidos de maneira progressiva, permitindo assim o acesso à leitura considerada *per se*, a chamada leitura expressiva. No entanto, para alguns, há de se considerar ainda uma quarta etapa: a da leitura silenciosa como uma leitura “oralizada” ou “subvocalizada” que, ao ser dominada pelo leitor, é interiorizada. Em contrapartida, para outros, a leitura silenciosa precede a leitura em voz alta, pois é uma atividade processadora de sentido, sem emissão sonora, mas necessária à essa modalidade.

No entanto, Bajard (2014) considera as terminologias e conceitualizações de “leitura silenciosa” e “leitura em voz alta” problemáticas. Do ponto de vista lexical, a primeira deveria ser chamada de leitura em voz baixa, uma vez que se trata de uma oposição à leitura em voz alta, esta, mostrando-se também insatisfatória como nomenclatura. Isso porque, enquanto a primeira se caracteriza como uma atividade pessoal e íntima, a segunda é sonora e comunicativa. Por isso, tomando como base as conceitualizações

de Eveline Charmeux (1987), Bajard (2014) argumenta que a leitura em voz alta não deveria nem ao menos ser considerada uma leitura, mas sim uma “transmissão” ou “emissão” sonora de um texto.

Seguindo por essa linha de raciocínio, Bajard (2014) propõe, então, o emprego das seguintes categorias: *oralizar*, para a atividade de identificação das palavras através da voz, ou seja, a decifração; *ler*, para a atividade silenciosa de construção de sentido de sentido a partir de um significante gráfico; e *dizer* para a atividade de comunicação vocal de um texto preexistente, sendo o produtor do texto “dito” chamado de *locutor*, apesar da ambiguidade com o profissional do rádio, pois o termo ideal, *ditador*, remete a um outro significante, ainda mais longe do ideal.

Em sua visão, o *dizer* recorre a uma grande variedade de linguagens para além dos signos linguísticos, como o olhar, os gestos, os figurinos e até mesmo o cenário. Mas quando o *dizer* se dá por meios puramente sonoros, como é o caso dos audiolivros, ele considera, a partir dos estudos de Anne Ubersfeld (1978), determinados signos acústicos, como as “[...] inflexões, suas pausas, seu timbre, seus suspiros, sua musicalidade, a voz, às vezes acompanhada de sonoplastia, atrai outras linguagens junto ao texto linguístico, produzindo uma obra nova” (Bajard, 2014, p. 92).

Entretanto, há situações em que o *dizer* pode ser neutralizado, como é o caso da leitura *recto-tono*, geralmente realizada em conventos e seminários, em que não há inflexões na voz, sendo as pausas as únicas marcas sintáticas. Embora na perspectiva do autor essa neutralidade seja possível, acreditamos que a totalidade dela não exista, visto que todo ato comunicacional traz uma intenção a qual é refletida, de alguma maneira, na fala humana, ainda que sutilmente.

As ideias de Bajard (2014) em conjunto com as de Zumthor (2018) nos permitem pensar que, apesar da ausência do corpo, ainda assim é possível perceber a performance pela escuta, ainda que ausente o elemento visual, por meio da prosódia. A Alexa, no entanto, não possui um corpo biológico para reagir ao que está “lendo”, que dizer, sua vocalização não tem conexão com as reações de um corpo àquilo que lê. Logo, por meio dos pressupostos teóricos de Zumthor (2018) e Bajard (2014), a Alexa não *performatiza* ou *diz*. Todavia, há uma *performance* envolvida, mas não a que entendemos dentro dos parâmetros humanos, mas, sim, maquínicos (de uma inteligência *artificial*) — questão essa que precisa ser debatida em virtude da rápida evolução de sistemas como esse, cada vez mais próximos das habilidades humanas. Não encontramos estudos para aprofundar essa questão.

Apesar disso, a complexidade dessa questão nos possibilita pensar em uma última relação: da consciência e da inteligência. Santaella, (2022), em seus estudos sobre inteligências artificiais, argumenta que consciência se trata de um conceito muito abrangente e, por isso, sua definição demanda a visão particular de cada filósofo ou cientista da cognição que buscou conceituá-la. Ainda que a autora apresente diferentes concepções, por estarmos tratando de leitura, a consciência é aqui entendida como a capacidade de estar no mundo e se relacionar com ele por meio de vivências pessoais. De qualquer forma, como bem explica a autora, a consciência se

trata de um “fenômeno interior, e por ser interior, é difícil encontrar formas de compartilhamento entre consciências, tanto quanto é difícil saber se outros animais também a têm” (Santaella, 2022, p. 3790). Ou seja, quando falamos de consciência, estamos falando a partir de nossa perspectiva, a humana, porque não somos capazes de alcançar outras consciências.

Embora — até o momento — as I.As., de maneira geral, não tenham essa consciência, como nós, humanos, a entendemos, o mesmo não se pode afirmar no que diz respeito à inteligência. Isso porque, na área das Ciências da Computação, o objetivo geral das I.A. como a Alexa é o de fazer com que elas emulem os processos de inteligência humana em sistemas computacionais a partir de modelos baseados em conhecimento e aprendizado, com o intuito de auxiliarem na melhoria dos processos de tomada de decisão de maneira cada vez mais eficiente, como afirmam os autores Stuart J. Russel e Peter Norvig (2004).

No contexto atual, é possível observamos um aumento cada vez maior de aplicações com os chamados *chatbots* (agentes conversacionais), como *smart bots*, entidades de conversação artificial e assistentes virtuais, a exemplo da própria Alexa. Essa nomenclatura é escolhida para esses tipos de sistemas porque cada interação se dá a partir de um conjunto de perguntas e respostas de conversação entre um agente virtual conversacional e um usuário humano. De modo progressivo, essa troca comunicacional faz com que esses agentes virtuais se especializem para compreender a linguagem e melhorem cada vez mais o processo de interação (ADAMOPOULOU, MOUSSIADES, 2020).

A aprendizagem de sistemas do tipo *chatbot* pode ser dividida em quatro modos: 1) aprendizado supervisionado: recebe dados de entrada (exemplos) associados a dados de saída (resultado), previamente rotulados, utilizados para o treinamento, possibilitando o reconhecimento de novos padrões; 2) aprendizado não-supervisionado: recebe apenas dados de entrada (exemplos), com o objetivo de realizar agrupamento de padrões similares (*clusterização*); 3) aprendizado por reforço: não há um conjunto de treinamento associado, ou exemplos rotulados, são algoritmos que trabalham por recompensas ou punições em tarefas realizadas; e, 4) (aquele que nos interessa particularmente por ser o tipo de aprendizado da Alexa) aprendizado profundo (*deep learning*): baseado em representações de dados, com várias camadas de processamento, podendo utilizar algoritmos de aprendizagem supervisionada e não-supervisionada nas diferentes camadas (Mitchell, 1997; Rezende, 2003).

O *Deep Learning* faz parte dos métodos de aprendizado de máquina que utilizam as Redes Neurais Artificiais como modelo de representação, permitindo ao sistema computacional descobrir, de modo automático, representações necessárias para a detecção de características ou classificação de dados (Lecun, Bengio & Hinton, 2015). Assim, cada usuário da Alexa terá uma assistente diferente, pois seu aprendizado dependerá dos dados fornecidos por cada um de acordo com os seus hábitos de consumo.

Entre as funcionalidades da Alexa, a que nos interessa é a do TTS, aquela por meio da qual ela realiza a “leitura” de e-books da biblioteca Kindle do usuário. Ainda que as assistentes aprendam com as rotinas e com a interação e mesmo que nem todas as pessoas utilizam essa função TTS da Alexa, ela está integrada em todos os dispositivos. Levando em conta todos esses elementos em torno da escuta, da leitura vocalizada, da performance e dos dispositivos I.A., passamos para a forma como Elleström (2010) compreende os fenômenos intermidiais para ser possível definirmos o que é a “leitura” da Alexa.

## **CRUZANDO FRONTEIRAS: OS FENÔMENOS INTERMIDIAIS**

Os fenômenos intermidiais considerados por Elleström (2021) em sua teoria tem por base duas perspectivas de análise. A primeira, a sincrônica, também denominada de heteromidialidade, se encontra fundamentada nos conceitos de Jørgen Bruhn (2010), que por vez toma as ideias de Mitchell (1994) e do próprio teórico (Elleström, 2010). Ao levar em consideração as características das mídias em um determinado momento, a heteromidialidade, também chamada de Integração de Mídia, considera a combinação e a integração entre elas em sua sincronidade para a formação de um tipo ou de um produto de mídia. É exemplo dessa combinação o modo como um tipo de teatro pode integrar um telão mostrando um filme em sua composição durante o espetáculo.

Ao buscar entender as relações entre um tipo de mídia anterior e um tipo de mídia subsequente — do livro ao audiolivro —, nosso interesse se concentra na segunda perspectiva, a diacrônica. Também chamada de transmidialidade, a perspectiva diacrônica é compreendida de duas maneiras: na primeira, a da transmidialidade intramidial, é considerada a transferência de produtos entre tipos de mídias básicas semelhantes, ou melhor dizendo, entre tipos de mídias que apresentam modos iguais ou muito similares. No segundo caso, a transmidialidade intermidial, esses modos básicos são distintos.

Elleström (2021) define “modos” como traços das mídias, os quais são reunidos em quatro categorias básicas inerentes a todas elas, chamadas de modalidades. Bruhn e Schirmacher (2022) explicam que elas são responsáveis por dar forma aos produtos ou tipos de mídias ainda que “desapareçam” no ato comunicacional, quando o que importa é que está sendo comunicado: há um objeto material ou um fenômeno físico de comunicação (sólido, líquido, gasoso e/ou plasmático ou, ainda, orgânico e inorgânico), que percebemos através dos nossos sentidos (olfato, paladar, visão, tato e audição) a partir também da sua composição espaçotemporal (altura, largura, profundidade e tempo), interagindo conosco. Essas três modalidades e seus respectivos modos são pré-semióticas, dizem respeito à mediação no ato de comunicação: a material, a sensorial e a espaçotemporal.

A quarta modalidade das mídias é a semiótica. No momento em que percebemos sensorialmente que há algo sendo comunicado e que precisamos interpretar, ativamos diferentes representações, as quais Elleström (2021) compreende por meio da

tricotomia peirceana, porém com outras nomenclaturas para cada um de seus modos: para o ícone, a ilustração; para o índice, a indicação; para o símbolo, a descrição.

Se os modos das modalidades material, sensorial, espaçotemporal e semiótica forem suficientes para que seja possível categorizar determinados tipos de mídias, elas passam a ser entendidas como sendo básicas, a exemplo de um cartaz com palavras escritas, que ao olhar sabemos se tratar de um cartaz. Mas, quando a análise dessas modalidades não basta — é um cartaz publicitário ou uma propaganda governamental? —, é preciso levar em conta ainda outros dois aspectos qualificadores, que são aqueles que utilizamos para compreender as diferenças entre o audiolivro, o programa de rádio e o programa de podcast em um estudo anterior.

O primeiro é o aspecto qualificador contextual, que diz respeito às práticas discursivas, convenções históricas e localização geográfica, como a história das mídias. Já o segundo é o aspecto qualificador operacional, que engloba as funções comunicativas das mídias que são esperadas ou presumidas, nos permitindo dizer que se trata de um cartaz de propaganda governamental, um anúncio publicitário ou um aviso de condomínio.

Como neste estudo estamos trabalhando com dois tipos de mídias muito diferentes entre si, visto que o livro é fundamentalmente visual e o audiolivro e a *leitura* realizada pela Alexa são fundamentalmente sonoras, podemos pensar em ambas enquanto tipos de mídias básicas neste momento. É analisando suas quatro modalidades que podemos concluir que tratamos de tipos de mídias com modos diferentes sobretudo no aspecto da mediação.

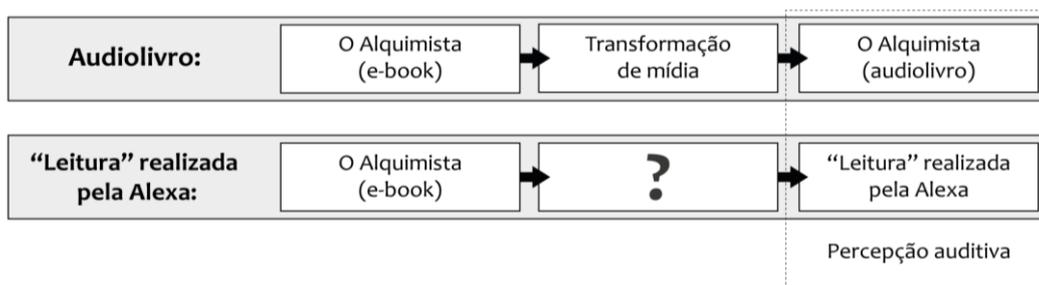
Por essa razão, o fenômeno que nos interessa é o da transmidialidade intermidial, também chamado de transformação de mídia. Ele diz respeito à transferência e transformação de um produto de mídia a, justamente, tipos de mídias distintas. Neste ponto, é importante salientar que, na visão de Elleström (2021), não há transferência sem que haja uma transformação, pois isso sempre ocorre ainda que seja pouco perceptível. Logo, não é possível “adaptar” sem transformar. São duas as transmediações que nos interessam: entre uma mídia verbal que se transforma em audiobook e entre uma mídia verbal que se transforma em “leitura vocalizada”.

Assim, para a melhor compreensão dos nossos próximos passos, é necessário idealizarmos um fluxograma bastante simplificado (Figura 1) pensando no paralelo entre o audiolivro e a “leitura” realizada pela Alexa. De um lado, um livro passa por um processo de transformação de mídia para se tornar um audiolivro: roteirização, gravação e regravação, montagem, edição, finalização e distribuição. Esse audiobook é acessado através de um dispositivo eletrônico como um computador ou celular e escutado pelo usuário. De outro, há o livro em formato digital, o qual também é midiado por um dispositivo eletrônico através do aplicativo Kindle. Por meio do comando de voz do usuário, a Alexa “lê” o conteúdo do livro digital para os usuários quando é solicitada. Apesar das diferenças, ambos fenômenos acabam sendo percebidos da mesma maneira: pela audição como principal sentido. Neste caso, como

até aqui não é possível afirmar o tipo de fenômeno, deixamos uma lacuna que será descoberta nas páginas seguintes.

Com isso em mente, no momento em que o livro digital de *O Alquimista* (Coelho, 2017) é realizado pelas referidas mídias técnicas de exposição externas, temos: o plasmático ou inorgânico como modo da modalidade material; a largura, a altura da distribuição da narrativa na página relacionada ao tempo dedicado à leitura como modos da modalidade espaçotemporal, sendo que a profundidade é desconsiderada neste caso, pois apesar de as telas terem suas próprias espessuras, isso não importa à leitura; a visão sobre as palavras e o tato, no virar de páginas, como modos da modalidade sensorial.

**Figura 1.** Fluxograma simplificado do paralelo proposto entre o audiolivro e a “leitura” da Alexa.



Já na modalidade semiótica podemos definir a descrição (símbolo) como principal modo, visto que o livro digital é composto por palavras cuja representação é convencional. Podemos considerar, ainda, os modos indicação (índice) e a ilustração (ícone) que emergem da secundidade, na relação entre a história lida — *O Alquimista* (COELHO, 2017) — e o mundo que ela representa. De um lado, a história cria um universo semelhante ao nosso; de outro a verossimilhança da narrativa se dá enquanto ela se “indicializa”, ou seja, enquanto a representação pelas palavras pode apresentar algo novo, mas reconhecível, alcançando a qualidade factual da experiência. No entanto, em primeiridade, na relação entre objeto e representamen, a descrição (símbolo) se sobrepõe aos outros modos de representação no ato da percepção (Quadro 1).

**Quadro 1.** Os modos do tipo de mídia e-book de *O Alquimista* (Coelho, 2017).

Produto de Mídia	Mídia técnica		Modalidades				Tipo de Mídia
	Hardware	Software	Material	Espaçotemporal	Sensorial	Semiótica	
<i>O Alquimista</i> (2017)	E-reader Kindle ou computador	Kindle Software	Inorgânico Plasmático	Largura Altura Tempo	Visão Tato	Descrição Ilustração Indicação	Livro digital

Em contrapartida, no audiolivro de *O Alquimista* (COELHO, 2021) narrado pela atriz Beth Goulart, o qual está disponível no Audible e pode ser midiado por meio dos alto-

falantes inteligentes Echo Dot, temos: a onda sonora como modalidade material, que por se tratar de um fenômeno físico, independentemente da biologia de sua fonte, é sempre orgânica; o tempo de exibição do áudio como um modo da modalidade espaçotemporal; e a audição como um modo da modalidade sensorial. Vejamos que as três modalidades pré-semióticas apresentam diferenças.

Da mesma maneira como acontece na modalidade semiótica do livro digital, a escuta semântica nos permite definir como modo principal a descrição (símbolo), já que se tratam de palavras ditas em português. Além disso, por meio da escuta reduzida que, como visto, se trata de um tipo de escuta puramente estético, é possível observarmos algumas convenções relacionadas às características acústicas da voz, como entonações, emoções e as próprias relações sintáticas-semânticas.

Uma leitura mais pausada, acompanhada de um volume de voz baixo, por exemplo, transmite a sensação de calma ou tranquilidade, enquanto uma leitura realizada mais rápida, com o volume de voz alto, traz com ela a sensação de euforia ou ansiedade — sentimentos esses que são potencializados pelos sons organizados e sons não verbais que os acompanham. No caso do audiolivro de *O Alquimista* (COELHO, 2021), é possível interpretarmos o dizer de Beth Goulart como uma transmissão ou emissão sonora carregada de inflexões, que vão variando de frases a frases entre uma leitura acelerada e lenta de acordo com a sua intenção, ou a de quem a dirigiu, com o objetivo de sensibilizar quem escuta.

Apesar de considerarmos a descrição (símbolo) como principal modo, também é possível encontrarmos os modos indicação (índice) e a ilustração (ícone) por diferentes motivos, dos quais citamos alguns. Por meio da escuta causal, por exemplo, é possível percebermos que a voz de Beth Goulart é indicial no sentido em que se trata de uma voz advinda de uma fonte humana e feminina fora do âmbito da representação, pois existe no mundo externo, e passível de ser reconhecida como pertencente a ela – os créditos do audiolivro nos permitem reconhecer o nome da atriz cuja voz é o índice da sua pessoa.

A presença de efeitos sonoros advindos de animais, como o balir dos ovinos — o “méééé!” — integrados ao dizer de Beth Goulart, também evocam a indicialidade, uma vez que se tratam de seres não humanos que possuem uma presença igualmente fora do âmbito da representação. Já outros tipos de efeitos sonoros presentes no audiolivro, como farfalhar de folhas para ilustrar o dizer que o personagem está lendo um livro para o rebanho, ou o bater do cajado no solo seco trazem consigo a iconicidade. Afinal, pela ausência do sentido da visão, o som remete ao ato ou ação dita, mas isso não significa que tenha vindo de fato desses objetos devido à capacidade da sonoplastia de imitar sons que se assemelham a partir de outras origens sonoras que muito se assemelha entre si.

Dessa forma, podemos estabelecer as seguintes modalidades ao audiolivro em relação ao e-book, conforme o Quadro 2.

**Quadro 2.** Comparação entre os modos do livro digital e do audiolivro de O Alquimista (Coelho, 2017; 2021).

Produto de Mídia	Modalidades						Tipo de Mídia
	Mídia técnica		Material	Espaçotemporal	Sensorial	Semiótica	
	Hardware	Software					
O Alquimista (2017)	E-reader Kindle ou computador	Kindle Software	Inorgânico Plasmático	Largura Altura Tempo	Visão Tato	Descrição Ilustração Indicação	Livro digital
<b>TRANSFORMAÇÃO DE MÍDIA</b>							
O Alquimista (2021)	Echo Dot	Audible	Ondas sonoras (orgânico)	Tempo	Audição	Descrição Ilustração Indicação	Audiolivro

Em vista de tais características, podemos definir que a mudança de um tipo de mídia a outro se dá nas modalidades pré-semióticas, fundamentalmente, na modalidade material e, neste caso, na mídia técnica de exposição externa que a realiza. Essa mudança significativa, no entanto, faz com as outras modalidades também sejam afetadas, incluindo aí os modos semióticos. Os efeitos da transmídiação na modalidade semiótica aparecem na intencionalidade que há por trás do “dizer”, na concepção de Bajard (2014) e Zumthor (2018), uma vez que, como visto anteriormente, o ato da decodificação ou decifração é neutro — já o dizer ou a performance, não. Junto aos sons organizados e os sons não verbais, podemos dizer que Beth Goulart interpreta o que lê e, assim, traz essa representação no modo como diz ou performatiza o texto, e esses significados dados à leitura são percebidos durante a escuta através dos mais variados elementos acústicos (pausas, diferenças no tom de voz, entre outros).

Por essa razão, há no audiolivro uma intenção genuína de transferência de um tipo de mídia a outro para oferecer um novo produto. Nesse cenário, se dermos o mesmo texto a uma outra pessoa, ou até mesmo solicitar que Beth Goulart realize uma nova leitura, é muito provável que haja a inclusão de novas representações, pois não é possível reproduzir com exatidão as mesmas condições como da primeira vez. Sendo assim, estamos tratando de uma transformação de mídia, uma transmídiação, em todos os traços das mídias. Vejamos que essa transmídiação pode ser equiparada àquela realizada pela adaptação de uma obra literária para o cinema, pois ambas exigem, em diferentes graus, a interpretação da história para sua transferência. Por outro lado, a I.A. não faz essa interpretação da história.

### A “LEITURA” DA ALEXA: UMA TRANSFORMAÇÃO DE MÍDIA?

A “leitura” da Alexa, quando comparada ao audiolivro, apresenta semelhanças nos modos das modalidades material, espaçotemporal e sensorial, e diferenças com relação à mídia técnica de exposição e na modalidade semiótica. A mídia técnica de

exposição externa (hardware) permanece a mesma e, ainda que o sistema (software) seja outro – o a da própria assistente virtual –, os modos pré-semióticos da “leitura” pela I. A. são os mesmos, baseados em onda sonora.

Ao contrário de Beth Goulart, e de qualquer outro locutor humano de um audiolivro que, obviamente, existe no mundo “real”, não é possível considerarmos a indicação (índice) — a não ser no campo da representação. A Alexa existe, mas não corresponde a uma entidade viva e, sim, à integração de subsistemas que foram programados e que, juntos, lhe dão a forma de uma entidade virtual composta por zeros e uns. Em comparação, Beth Goulart nunca leria completamente igual a mesma história, já a Alexa sempre lê da mesma maneira. Isso, no entanto, não a impede de ser programada para tentar emular sinteticamente o dizer ou a vocalização humana performatizada, visto que parece haver uma tentativa de torná-la cada vez mais natural.

Por isso, para considerar a indicação e a ilustração, é preciso levar em conta (muitas) ressalvas, pois ela não compreende e interpreta a mídia verbal escrita como Beth Goulart o faz. Além disso, a “leitura” da Alexa não conta com a integração de sons organizados e sons não verbais como o audiolivro. Há, assim, somente a sua “leitura”, sem nenhum outro recurso acústico que aja sobre o sentido das palavras.

Sendo assim, podemos afirmar, partindo da perspectiva de Zumthor (2018), que não há performance na Alexa, pois não há um corpo orgânico para reagir ao texto, rompendo assim com o fazer poético. Já pela perspectiva de Bajard (2014), podemos entender que o que a Alexa realiza é o processo de *oralizar*, ou seja, consiste na identificação das palavras através da decifração por meio da voz, sem ser possível atingir o dizer pela sua falta de consciência.

Em suma, o que a Alexa faz é reproduzir uma sequência de palavras e frases que, ao serem dispostas de acordo com o conteúdo do livro, somente fazem sentido aos ouvintes por meio da escuta semântica, uma vez que a assistente virtual da Amazon não possui uma consciência que se iguale ou supere a mente humana. E, por consequência, é incapaz de interpretar o conteúdo do que está “lendo”.

Dessa forma, ela não realiza a ação de *ler*, descrita por Bajard (2014) como uma atividade silenciosa de construção de sentido a partir do significante gráfico, mas consegue *oralizar*, isto é, identificar as palavras através da voz. Podemos então compará-la muito superficialmente a uma pessoa em processo de alfabetização que consegue decifrar ou decodificar um texto, mas sem conseguir entender o conteúdo do que está lendo.

Assim, diferentemente do audiolivro narrado por Beth Goulart, a Alexa não recria a obra ao seu modo, pois não a interpreta para vocalizar. Por isso, quando nos referimos à sua “leitura”, ou devemos mencionar entre aspas, pois há vários processos inerentes à leitura em si que ela (ainda) não é capaz de realizar, ou devemos chamar de *oralização* (BAJARD, 2014) — ou simplesmente podemos chamar de decifração ou decodificação.

Podemos definir, assim, as seguintes modalidades para a “leitura” da Alexa, conforme Quadro 3.

**Quadro 3.** Modos da oralização (Bajard, 2014) realizada pela Alexa.

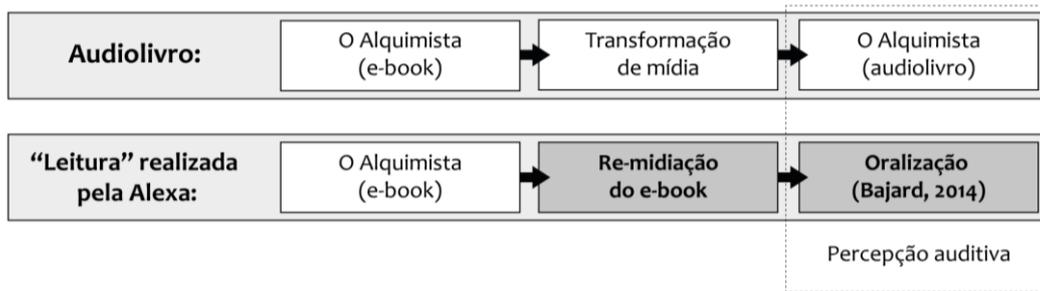
Produto de Mídia	Modalidades						Tipo de Mídia
	Mídia técnica		Material	Espaçotemporal	Sensorial	Semiótica	
	Hardware	Software					
O Alquimista (2017)	Echo Dot	Alexa	Ondas sonoras (orgânico)	Tempo	Audição	Descrição	Decodificação ou decifração

Assim, em um âmbito geral, a Alexa acaba por não criar um novo tipo de mídia como é o caso do audiolivro que, como temos comumente em mente, é lido por um ser humano, gravado, editado, distribuído, comercializado. No entanto, é passível de ser reproduzido quantas vezes forem necessárias da mesma forma, com as mesmas condições como se fosse da primeira vez, pois ela não traz em sua *oralização* (BAJARD, 2014) as inflexões e alterações no volume de voz como um ser humano faz — ao menos, não até a finalização desse estudo. Logo, por mais que os audiolivros e a *oralização* (BAJARD, 2014) acabem convergindo para uma percepção muito semelhante, somente é possível considerarmos o primeiro enquanto uma transformação de mídia, quer dizer, enquanto a transmídiação de uma mídia verbal para uma mídia sonora.

Para nós, a *oralização* (BAJARD, 2014) nos parece um caso não previsto por Elleström (2021) em sua teoria, por isso chamaremos esse fenômeno de “re-mídiação”. Para Elleström (2021), uma segunda mídiação de um produto de mídia é uma transmídiação, fenômeno que ele destaca como “de transformação”. Nossa intenção com o termo “re-mídiação” — mediar novamente — é justamente distinguir do processo de transmídiação, que visa à mudança da mídia e que constitui por um processo de adaptação. A “re-mídiação” é uma re-exibição ou uma re-realização das qualidades material, espaçotemporal, sensorial e potencialmente semióticas de um mesmo produto de mídia através de outro meio técnico. É importante pontuarmos que esse fenômeno de “re-mídiação” não é aquele da remediação (remediation) descrita por Bolter e Grusin em *Remediation: understanding new media* (2000), que diz respeito aos modos como as mídias re-elaboram mídias anteriores a partir das novas tecnologias. O significado de remediação tem uma conotação diacrônica de transformação de tipos de mídias qualificadas, enquanto aqui tratamos de fenômenos específicos de um único produto de mídia. Dessa forma, podemos resumir o fenômeno da seguinte forma: o usuário solicita à Alexa a *oralização* (BAJARD, 2014) do livro digital por meio do alto-falante inteligente Echo, o qual é “re-midiado” pela assistente virtual.

Assim, ao retomarmos ao fluxograma simplificado, podemos concluí-lo da seguinte maneira (Figura 2):

Figura 2. Fluxograma simplificado do paralelo proposto.



Assim, chama-nos a atenção a possibilidade de pensarmos em uma terceira forma de transferência de um mesmo produto de mídia para tipos de mídias diferentes sem que haja, necessariamente, uma transformação propriamente dita como Elleström (2021) propõe. Mas, para isso, é necessário aprofundarmos tal direcionamento em pesquisas futuras, verificando se o mesmo fenômeno ocorre com outros tipos de mídias.

## CONSIDERAÇÕES FINAIS

Ao longo desta discussão, pudemos demonstrar que os Estudos em Intermedialidade se caracterizam como um modo de observação e interpretação das mídias por meio da desconstrução de seus elementos e da comparação entre elas. Assim, de maneira geral, essa perspectiva nos possibilita distinguir e categorizar as mídias e os fenômenos que a evocam, bem como analisar e compreender suas respectivas relações ao se apropriar de diferentes perspectivas analíticas, como as da semiótica.

Visando, então, compreender se a “leitura” realizada pela Alexa se equipara à uma transmediação de um livro, resultando em um audiolivro, nos deparamos, por meio da proposta de Lars Elleström (2021) e trazendo ainda outros autores que tratam do tema, com um fenômeno que somente é possível graças à popularização das novas formas de comunicação digitais. A intermedialidade, como metodologia de análise, permite a comparação entre tipos e produtos de mídia e, desconstruindo-os a partir da observação de suas modalidades, contribui para a compreensão de suas semelhanças e diferenças.

Incluir o conceito de “re-mediação” evoca entender fenômenos que têm sido comuns hodiernamente, por exemplo, a passagem das imagens em movimento do SD (*Standard Definition*) para HD (*High Definition*) ou ainda novas edições de um mesmo produto de mídia, às vezes por atualização de softwares e aplicativos.

Até aqui, a “re-mediação” fica, assim, restrita a esses fenômenos de re-exibição das configurações sensoriais por uma nova mídia técnica de exposição, que são percebidos pelos receptores humanos de sentido em uma nova situação comunicativa. Em contrapartida, a transmediação fica restrita a re(a)presentação, quer dizer, a interpretação e a transformação de um produto de mídia em outro, que requer processos complexos de adaptação.

Especificamente em relação aos audiolivros, essa transformação pode ser feita de diferentes maneiras a partir da performance vocal do agente da voz, pois há aqueles cujo intuito é apenas decodificar as palavras em voz alta, enquanto outros são altamente interpretativos. Mas, em vista da necessidade de um aprofundamento teórico, abordaremos tal diferenciação e, possivelmente, classificação, para um estudo subsequente a esse.

Certamente a comparação entre o audiolivro e a “re-midiação” pela Alexa do e-book de *O Alquimista* (Coelho, 2017 [1988]; 2021), que nos levou a entender ambas como fenômenos diferentes, senão midiáticos, pelo menos semióticos, leva em conta as condições atuais da I. A. ainda ineficiente na emulação da performance humana ou incapaz do dizer ao modo de Bajard (2014). É bem possível que essas diferenças semióticas, dadas pela ausência ou presença de um corpo que interpreta e age na enunciação, sejam extintas brevemente.

## AGRADECIMENTOS

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), que por meio do Programa de Suporte à Pós-Graduação de Instituições Comunitárias de Educação Superior (PROSUC), viabilizou essa pesquisa sob forma de bolsas de estudo para os autores.

## FINANCIAMENTO

Este trabalho de pesquisa foi financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (processo número: 304566/2021-7).

## REFERÊNCIAS

ADAMOPOULOU, E.; MOUSSIADES, L. Chatbots: History, technology, and applications. Em: *Machine Learning with Applications*. Vol 2. 2020.

BAJARD, Élie. *Ler e dizer: compreensão e comunicação do texto escrito*. São Paulo, SP: Cortez, 2014.

BOLTER, Jay David. GRUSIN, Richard. *Remediation: Understanding New Media*. Cambridge (MA): MIT Press, 2000.

BROCH, José Carlos. *O conceito de affordance como estratégia generativa no design de produtos orientado para a versatilidade*. [Em linha] Dissertação (Mestrado em Design e Tecnologia). Porto Alegre, RS: Universidade Federal do Rio Grande do Sul. [Acesso em 22 janeiro 2022]. Disponível em:

<https://www.lume.ufrgs.br/bitstream/handle/10183/25510/000752864.pdf>

- BRUHN, Jørgen Bruhn; SCHIRRMACHER, Beate. *Intermedial studies*. Em: BRUHN, Jørgen Bruhn; SCHIRRMACHER, Beate. *Intermedial studies: an introduction to meaning across media*. Nova York, NY: Routledge, 2022.
- CHARMEUX, Eveline. *Apprendre à lire: échec à l'échec*. Paris: Milan, 1987.
- CHION, Michel. *The three listening modes*. Em: STERNE, Jonathan. *The Sound Studies Reader*. Nova York, NY: Routledge, 2012.
- COELHO, Paulo. *O Alquimista*. Ledor: Beth Goulart. São Paulo, SP: Paralela, 2021. Audiolivro.
- COELHO, Paulo. *O Alquimista*. São Paulo, SP: Paralela, 2017. E-book.
- ELLESTRÖM, Lars. *As modalidades das mídias II: um modelo expandido para compreender as relações intermidiais*. Tradução: Beatriz Alves Cerveira, Júlia de Oliveira Rodrigues e Juliana de Oliveira Schaidhauer. Porto Alegre: EDIPUCRS, 2021.
- ELLESTRÖM, Lars. *The Modalities of Media: A Model for Understanding Intermedial Relations*. Em: ELLESTRÖM, Lars (ed.). *Media Borders, Multimodality and Intermediality*. Basingstoke, Inglaterra: Palgrave Macmillan, 2010. p. 11-48.
- GOLD, Ben. MORGAN, Nelson. ELLIS, Dan. *Speech and audio signal processing: processing and perception of speech and music*. Nova Jersey, NJ: Wiley, 2011.
- HUANG, Xuedong; ACERO, Alex; HON, Hsiao-Wuen. *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. Nova Jersey, NJ: Prentice Hall, 2001.
- LECUN, Y., BENGIO, Y., HINTON, G. (2015). *Deep learning*. Em: *Nature*, 521(7553), p 436-444.
- MITCHELL, T. M. (1997). *Machine learning*. Burr Ridge, IL: McGraw Hill, 45(37), p. 870-877.
- MITCHELL, William John Thomas. *Picture Theory: Essays on Verbal and Visual Representation*. Chicago: University of Chicago Press, 1994.
- RABINER, L. R.; SCHAFER, R. W. *Introduction to digital speech processing*. Nova Jersey, NJ: Prentice Hall, 2010.
- REZENDE, Solange Oliveira. *Sistemas Inteligentes – Fundamentos e Aplicações*. São Paulo: Manole, 2003.
- RUSSEL, Stuart J.; NORVIG, Peter. *Inteligência Artificial*. Rio de Janeiro: Elsevier, 2004.
- JENSEN, Signe Kjær; SALMOSE, Niklas. *Media and modalities – Film*. In: BRUHN, Jørgen; SCHIRRMACHER, Beate. (Aut.). *Intermedial studies: an introduction to meaning across media*. 1 ed. Nova York: Routledge, 2022. p. 28-41
- SANTAELLA, Lucia. *Comunicação Ubíqua: repercussões na cultura e na educação*. São Paulo: Paulos, 2013.

SANTAELLA, Lucia. *Matrizes da linguagem do pensamento: sonora visual verbal: aplicações na hipermídia*. 3ª Ed. São Paulo: Iluminuras, 2005.

SANTAELLA, Lucia. *Neo-humano: a sétima revolução do Sapiens*. São Paulo: Paulus, 2022. Edição do Kindle.

SONNENSCHNEIN, David. *Sound design: the expressive power of music, voice, and sound effects in cinema*. Michigan: Michael Wiese Productions, 2001.

TAYLOR, Paul. *Text-to-Speech Synthesis*. Nova York, NY: Cambridge University Press, 2009.

ZUMTHOR, Paul. *Performance, recepção, leitura*. Tradução: Jerusa Pires Ferreira e Suely Fenerich. São Paulo, SP: Ubu Editora, 2018.

ZUMTHOR, Paul. “A poesia e a voz.” In: ZUMTHOR, Paul. *Escritura e nomadismo: entrevistas e ensaios*. Cotia, SP: Ateliê Editorial, 2005.