**1. Introdução**

Reconhecimento de emoção em fala é uma linha de pesquisa dentro da área de Inteligência Artificial (IA), dada por tarefas de reconhecimento e classificação da reação afetiva de um indivíduo [H]. Este estudo sobre emoções, sua interpretação e sua representação no contexto computacional formam uma área de estudos denominada por Computação Afetiva [I].

Sejam transmitidas pelo rosto, corpo ou voz, as expressões de emoção são onipresentes. O sentido inferido das expressões está, em geral, substancialmente alinhado com o conteúdo afetivo expresso, sendo intuitivo sugerir que quanto mais forte o estado afetivo expresso mais nítido o sentido emocional inferido [K].

Antes mesmo de conhecer as palavras, ou de conseguir pronunciá-las corretamente, já empregamos a fala como uma forma primária de comunicação e expressão de emoções [A]. Independente da idade, os indivíduos da espécie humana expressam emoções comuns (e.g.: alegria, raiva, medo). Entretanto, idiomas distintos podem produzir diferenças na forma como essas emoções são expressas em matéria de tom de fala e voz [B] [C].

Ao nos comunicarmos utilizando a voz [D], além de decodificar e interpretar o valor presente na mensagem, também decodificamos e interpretamos outros elementos (e.g.: Entonação e ritmo) para compreender a mensagem de maneira completa. Não é difícil perceber que um "Bom dia!" sorridente e efusivo seria interpretado de forma quase que diametralmente oposta de um "Bom dia..." dito de forma lenta e com pouca energia. Portanto, reconhecemos a emoção na fala, e essa emoção é uma variável para interpretar o que foi dito.

A fala é a maneira mais popular de se comunicar com os outros na vida diária e é amplamente usada para expressão emocional [E]. Pode transportar dois tipos de informação: Informação literal e informação relativa [F]. A informação literal destaca o significado direto e informação relativa significa as mensagens implícitas, como a emoção contida na fala [F]. A fala é sempre uma fonte potencial do estado emocional de uma pessoa. No domínio do Aprendizado de Máquina (*Machine Learning*, ML), o Reconhecimento da Emoção da Fala (*Speech Emotion Recognition*, SER) é conhecido como a tarefa de determinar e classificar as características emocionais da fala. SER tem enfrentado consistentemente problemas desafiadores de ML devido à complexidade dos sinais de fala [F].

Emoções têm papel importante na comunicação humana. No contexto natural, utilizamos várias informações do ambiente para conseguirmos detectar emoções em suas expressões. Assim, é seguro afirmar que a fala é um dos elementos relevantes ao se tentar observar quais emoções estão sendo expressadas por um interlocutor. Nesse contexto, existem trabalhos que afirmam que emoções são expressas diferentemente pela fala humana e que ouvintes são capazes de corretamente inferir o estado emocional de um interlocutor apenas com a informação da voz [J].

Modelos estatísticos e de ML, vêm sendo um dos possíveis caminhos para realizar tarefas de reconhecimento de emoção na fala desde o final do século XX [L] [M]. Embora, conseguir identificar o estado emocional de um sujeito não seja uma tarefa trivial, pois demanda uma capacidade de percepção apurada. Em seu contexto original, os interlocutores utilizam várias informações visuais, auditivas, semânticas e metalinguísticas [N] para determinar qual emoção a fala de uma pessoa invoca, o que torna a tarefa bastante complexa e propensa a erros para o contexto da IA.

Ademais, existem trabalhos que demonstram que é possível inferir a emoção expressada em uma representação digital de uma fala, com diferentes técnicas de inteligência artificial [O] [P] [Q], bem como a intensidade de uma emoção na voz [K]. Assim, são confirmadas as suposições importantes para este trabalho:

1. É possível inferir o estado emocional de um interlocutor apenas com a informação da fala;
2. É possível solucionar a tarefa de reconhecimento de emoção na voz por técnicas de IA;
3. É possível inferir um valor para quantificar a intensidade de dada emoção na fala do interlocutor.

**1.1 Justificativa**

Pesquisando trabalhos científicos relacionados conseguimos encontrar diversas publicações. Artigos propondo modelos [T] e arquiteturas para SER já no ano de 2005, *reviews* [R] e *surveys* [S] comparando e detalhando tanto modelos de classificação quanto as bases de dados (*datasets*) utilizadas para treinar e validar os modelos das publicações.  
 Tornando a pesquisa mais específica, buscando por trabalhos que envolvam SER para a língua portuguesa (*PT-BR*), vamos perceber que o primeiro dataset em português brasileiro foi publicado em 2018: VERBO [V] [U].  
 Para a proposta deste trabalho, que é de tentar lidar não só com a emoção, mas tentar quantificar a intensidade da emoção do interlocutor falante de português, não foram encontrados nenhum trabalho relacionado e nenhuma base de dados - com anotações relativas a intensidade - no idioma desejado.  
 Assim, este trabalho se propõe a inovar, pesquisando técnicas para encontrar uma forma de inferir a intensidade da emoção em uma sentença falada. Entretanto, não partirá do zero, uma vez que já existem abordagens consolidadas para lidar com dados de forma semi-supervisionada e não supervisionada [W] que podem ser utilizadas com ponto de partida ou para estudo comparativo.  
 O resultado potencial deste trabalho poderia ser aplicado das seguintes formas, dentre outras:

1. Melhora na interpretação na entrada de dados (input) e da fidelidade das respostas (output) de assistentes virtuais;
2. Encaixar a solução numa arquitetura para canais de atendimento, vindo a fornecer uma melhor experiência do usuário (User Experience, UX);
3. Combinar a inferência da intensidade da emoção com outros tipos de solução de classificação de IA para obter resultados mais completos, complexos e fidedignos [X];
4. Criar um produto comercial (*white label*) para conversação mediada por IA [Y];
5. Tentar criar perfis comportamentais para funcionários e clientes, criando um emparelhamento otimizado das partes para melhorar a experiência, podendo ocasionar mais vendas de produtos financeiros;
6. Criando um novo dataset para futuras produções científicas.

**1.2 Alinhamento do tema às Estratégias Corporativa do BB**

Sabendo que uma das estratégias corporativas é ser o Banco com a plataforma de negócios e serviços mais relevantes para o cliente, proporcionando a melhor experiência, a proposta deste trabalho vai diretamente de encontro a essa ideia.

**1.3 Apresentação geral das principais questões a serem investigadas (pergunta da pesquisa)**

**1.4 Objetivos Específicos**

**1.5 Objetivo Geral**

**2. Referencial teórico**

**2.1 Fundamentação teórica do tema (incluir uma revisão dos principais conceitos que fundamentam o tema e priorizar a literatura atualizada – últimos 5 anos**

**3. Procedimentos metodológicos**

**3.1 Delineamento do estudo.**

**3.2 Participantes**

**3.3 Procedimentos de coleta de dados com definição de onde será realizada a pesquisa Banco do Brasil.**

**3.4 Instrumentos utilizados**

**3.5 Procedimentos de análise de dados.**

**3.6 Procedimentos éticos**

**3.7 Resultados esperados**

**3.8 Descrição dos resultados e das contribuições práticas da pesquisa, tanto para o funcionário, quanto para o Banco do Brasil.**

**5. Cronograma e orçamento**

**5.1 Identificação das principais etapas e atividades a serem realizadas durante a pesquisa, bem como de suas datas de conclusão**

**6. Referências bibliográficas**

[A] P. Lieberman, “The evolution of human speech: Its anatomical and neural bases” Current anthropology, vol. 48, no. 1, pp. 39–66, 2007.

[B] C. Kramsch, “Language and culture,” AILA review, vol. 27, no. 1, pp. 30–55, 2014.

[C] E. Sapir, Language: An introduction to the study of speech. Harcourt, Brace, 1921

[D] CASTRO, Sara de. "Elementos da comunicação"; Brasil Escola. Disponível em: https://brasilescola.uol.com.br/redacao/elementos-presentes-no-ato-comunicacao.htm. Acesso em 04 de novembro de 2022.

[E] Sahidullah, M.; Saha, G. Design, analysis and experimental evaluation of block based transformation in MFCC computation for

speaker recognition. Speech Commun. 2012, 54, 543–565.

[F] Garrido, M. The Feedforward Short-Time Fourier Transform. IEEE Trans. Circuits Syst. II Express Briefs 2016, 63, 868–872.

[G] Angadi, S.; Reddy, V.S. Hybrid deep network scheme for emotion recognition in speech. Int. J. Intell. Eng. Syst. 2019, 12, 59–67.

[H] Moataz El Ayadi, Mohamed S Kamel e Fakhri Karray. “Survey on speech emotion recognition: Features, classification schemes, and databases”. Em: Pattern Recognition 44.3 (2011), pp. 572–587

[I] Stuart Russell e Peter Norvig. “Artificial intelligence: a modern approach”.

[J] Klaus R Scherer. “Expression of emotion in voice and music”. Em: Journal of voice

9.3 (1995), pp. 235–248.

[K] Holz, N., Larrouy-Maestri, P. & Poeppel, D. The paradoxical role of emotional intensity in the perception of vocal affect. Sci Rep 11, 9663 (2021). https://doi.org/10.1038/s41598-021-88431-0

[L] Frank Dellaert, Thomas Polzin e Alex Waibel. “Recognizing emotion in speech”.

Em: Proceeding of Fourth International Conference on Spoken Language Processing.

ICSLP’96. Vol. 3. IEEE. 1996, pp. 1970–1973.

[M] Oh-Wook Kwon et al. “Emotion recognition by speech signals”. Em: Eighth European Conference on Speech Communication and Technology. 2003.

[N] Klaus R Scherer. “Expression of emotion in voice and music”. Em: Journal of voice

9.3 (1995), pp. 235–248.

[O] Yixiong Pan, Peipei Shen e Liping Shen. “Speech emotion recognition using support

vector machine”. Em: International Journal of Smart Home 6.2 (2012), pp. 101–

108.

[P] Kun Han, Dong Yu e Ivan Tashev. “Speech emotion recognition using deep neural

network and extreme learning machine”. Em: Fifteenth annual conference of the

international speech communication association. 2014.

[Q] Kun-Yi Huang et al. “Speech emotion recognition using deep neural network considering verbal and nonverbal speech sounds”. Em: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE.

2019, pp. 5866–5870.

[R] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar and T. Alhussain, "Speech Emotion Recognition Using Deep Learning Techniques: A Review," in IEEE Access, vol. 7, pp. 117327-117345, 2019, doi: 10.1109/ACCESS.2019.2936124.

[S] Abbaschian, B.J.; Sierra-Sosa, D.; Elmaghraby, A. Deep Learning Techniques for Speech Emotion Recognition, from Databases to Models. Sensors 2021, 21, 1249. https://doi.org/10.3390/s21041249

[T] Lin, Y.L.; Wei, G. Speech emotion recognition based on HMM and SVM. In Proceedings of the 2005 International Conference on Machine Learning and Cybernetics, Guangzhou, China, 18–21 August 2005; Volume 8, pp. 4898–4901

[U] TORRES NETO, José R. et al. VERBO: voice emotion recognition database in portuguese language. Journal of Computer Science, v. 14, n. 11, p. 1420-1430, 2018Tradução. Disponível em: http://dx.doi.org/10.3844/jcssp.2018.1420.1430. Acesso em: 04 nov. 2022.

[V] Neelakshi Joshi. "Brazilian Portuguese emotional speech corpus analysis". X Seminário em TI do PCI/CTI. 2021. Disponível em: https://www.gov.br/cti/pt-br/publicacoes/producao-cientifica/seminario-pci/xi\_seminario\_pci-2021/pdf/seminario-2021\_paper\_29.pdf. Acesso em 04 de novembro de 2022.

[W] S. E. Eskimez, Z. Duan and W. Heinzelman, "Unsupervised Learning Approach to Feature Analysis for Automatic Speech Emotion Recognition," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 5099-5103, doi: 10.1109/ICASSP.2018.8462685.

[X] Liu, Rui, et al. "Accurate Emotion Strength Assessment for Seen and Unseen Speech Based on Data-Driven Deep Learning." arXiv preprint arXiv:2206.07229 (2022).

[Y] Behavioral Signals. Disponível em: https://behavioralsignals.com/ai-mediated-conversations-case-study/. Acesso em 04 de novembro de 2022.

[Z]

[A1]

[B1]

[C1]

[D1]

[E1]

[F1]

[G1]

[H1]

[I1]

[J1]

[K1]

[L1]

[M1]

[N1]

[O1]

[P1]

[Q1]

[R1]

[S1]

[T1]

[U1]

[V1]

[W1]

[X1]

[Y1]

[Z1]

**7. Sugestão de livro para ajudar na elaboração do projeto**

Goodfellow et al. Deep Learning. An MIT Press book. 2016. Disponível em: https://www.deeplearningbook.org/