Detecção e Incorporação de Emoções na Tradução Automática para Libras: um Desenho de Pesquisa

Jenniffer Kelly R. de Araújo[1], **Manuella A. Lima**[1], **Diego L. R. da Silva**[1], **Daniel F. L. de Souza**[1], **Tiago M. U. de Araújo**[1]

[1]Centro de Informática, Universidade Federal da Paraíba, João Pessoa/PB, Brasil

{jenniffer.kelly, manuella.lima, diego.luis}@lavid.ufpb.br

{daniel,tiagomaritan}@lavid.ufpb.br

Abstract. *Machine translation into Brazilian Sign Language plays a crucial role in reducing barriers to information access for deaf people, especially through avatars such as VLibras. However, these tools remain limited by expressive neutrality, lacking a dynamic representation of affective Non-Manual Expressions. This work proposes a modular architecture that integrates automatic emotion detection in Portuguese text using LLMs to identify affective states and integrate them into the animation module. The research seek to improve the perception of naturalness and acceptability through a controlled experiment.*

Resumo. *A tradução automática para a Língua Brasileira de Sinais tem papel crucial na redução das barreiras de acesso à informação, especialmente por meio de avatares tal como no VLibras. Contudo, essas ferramentas permanecem limitadas pela neutralidade expressiva, carecendo de uma representação dinâmica das emoções através das Expressões Não Manuais. Este trabalho propõe uma arquitetura modular que integra um módulo de inferência emotiva no texto, baseado em LLMs, para identificar o estado emocional e incorporar essas informações emotivas ao processo de tradução. A pesquisa busca melhorar a percepção de naturalidade e a expressividade emotiva, contrastando o avatar neutro com uma versão emocional adaptada.*

1. Introdução

A Língua Brasileira de Sinais (Libras), enquanto língua de modalidade gestual-visual com gramática própria, é a primeira língua (L1) da comunidade surda no Brasil. No cenário da transformação digital, ferramentas de tradução automática como o VLibras são vitais para a acessibilidade; entretanto, esses sistemas ainda enfrentam limitações críticas na expressividade facial dos avatares. Essa neutralidade emocional compromete a fluidez comunicativa [Silva 2021], uma vez que as Expressões Não Manuais (ENMs) não são adornos, mas componentes gramaticais obrigatórios que transmitem intenção e carga emotiva [Wolfe and McDonald 2021].

A ausência de uma representação facial dinâmica resulta em uma sinalização robótica, elevando o esforço cognitivo do usuário e prejudicando a inteligibilidade da mensagem [Moraes et al. 2018]. Para superar essa rigidez e aproximar a tradução automática da competência comunicativa humana, este trabalho descreve o desenho de uma pesquisa e os resultados iniciais de uma arquitetura modular para inferência emotiva. A investigação

utiliza Grandes Modelos de Linguagem (LLMs) para identificar cinco estados emotivos (felicidade, tristeza, medo, raiva e surpresa) no texto de origem, associando esses rótulos a parâmetros de animação facial em um sistema procedural.

A principal contribuição deste estudo é a validação de uma arquitetura que automatiza a extração de estados emotivos via LLMs, incorporando-os dinamicamente ao processo de tradução para Libras. Ao sincronizar as ENMs com a sinalização manual de acordo com o contexto semântico do texto original, a abordagem mitiga as lacunas de expressividade das ferramentas atuais. Dessa forma, potencializa-se a experiência de acessibilidade para o usuário surdo, garantindo uma comunicação mais natural e eficaz.

2. Trabalhos Relacionados

A literatura sobre tradução automática para línguas de sinais tem evoluído significativamente na conversão de glosas e animação manual. Historicamente, pesquisas focaram na precisão lexical, como observado no sistema VLibras, que se consolidou no Brasil como a principal ferramenta de código aberto. Entretanto, o foco primordial no léxico manual resulta em uma neutralidade emocional que compromete a fluência humana. Para mitigar essa lacuna, [Saunders et al. 2020] utilizaram Progressive Transformers para buscar maior naturalidade na produção de sinais de ponta a ponta.

No contexto nacional, [Silva 2021] demonstrou que a incorporação de Expressões Não Manuais (ENMs) emotivas melhora significativamente o realismo e a aceitabilidade de avatares, embora tenha operado sob uma lógica de valência binária de sentimento (positivo/negativo). Essa necessidade de expressividade emotiva é reforçada pela ótica linguística de [Wolfe and McDonald 2021], que evidenciam que a neutralidade expressiva é um desafio global que distancia os sistemas da fluência observada em intérpretes humanos.

Recentemente, a transição para modelos de larga escala marcou o estado da arte. [Clemente 2024] observa que as ENMs são componentes gramaticais obrigatórios, e sua ausência eleva o esforço cognitivo do usuário. Por fim, [Fang et al. 2024] propõem o SignLLM, demonstrando que o uso de LLMs permite gerar gestos com qualidade superior ao tratar a língua de sinais como uma linguagem natural complexa. Este trabalho alinha-se a essa tendência, utilizando o DeepSeek-V3 para superar a rigidez de modelos lexicais tradicionais (como SVM ou BERT) e capturar nuances emotivas granulares.

3. Arquitetura de Detecção e Incorporação de Emoções

A arquitetura proposta (ver Figura 1) foi desenvolvida sob uma perspectiva modular, visando a integração transparente com sistemas de tradução para Libras já existentes. O fluxo de dados é estruturado em três camadas sequenciais que transformam o texto de entrada em parâmetros de animação facial para o avatar.

3.1. Pré-processamento e Gateway de API

O processo inicia-se com o recebimento do texto em português por meio de um Gateway de API desenvolvido em Node.js, responsável por orquestrar o fluxo entre os módulos da arquitetura. Nesta etapa, o sistema realiza normalização Unicode (NFC) e sanitização por meio de expressões regulares, garantindo a integridade do texto antes da inferência emotiva. O conteúdo é estruturado em objetos JSON, preservando metadados necessários para a sincronização posterior com o módulo de tradução lexical.

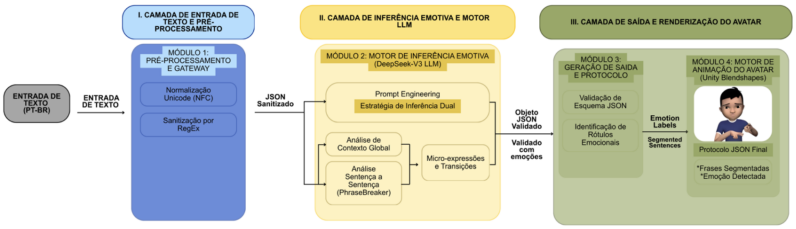

Figura 1. Arquitetura modular para detecção e incorporação de emoções

3.2. Motor de Inferência Emotiva via LLM

A detecção dos estados afetivos é realizada por um motor de inferência baseado no modelo DeepSeek-V3, acessado por meio de chamadas de API REST. A classificação emocional é conduzida via Prompt Engineering, dispensando treinamento supervisionado e permitindo a identificação contextual de nuances efetivas no texto. O módulo retorna uma categoria emocional entre seis classes predefinidas, que orienta a ativação das ENMs no avatar.

3.2.1. Estratégia de Prompt Engineering e Mitigação de Viés

A detecção via DeepSeek-V3 utiliza Role-Based Prompting e Few-Shot Chain-of-Thought para mitigar o "viés de negatividade"observado em modelos como GPT-4o-mini e Gemini 1.5 Flash. A estratégia opera em três níveis: persona especialista, taxonomia restrita a seis categorias e análise de subtexto para transições ambíguas. Para evitar a classificação de frustração como "Raiva", o prompt incorporou negativos específicos e ponderação contextual. Isso garante que termos como "pesado", em contextos de superação, sejam interpretados corretamente no escopo global.

3.2.2. Estratégia de Inferência Dual: Contexto Global e Micro-Expressões

A Estratégia Dual supera modelos tradicionais ao processar o texto em duas camadas simultâneas: o Contexto Global estabelece uma linha de base coerente com o tom predominante, enquanto a camada Sentença a Sentença utiliza o *PhraseBreaker* para identificar microexpressões e transições (ex: conjunções adversativas). Essa abordagem permite transições dinâmicas (ex: tristeza para felicidade), aproximando a performance do avatar à de um intérprete humano. Os rótulos são convertidos em *blendshapes* no Unity, sobrepostos aos sinais manuais. A integração garante que a carga emotiva seja fluida, preservando a gramática da Libras e reduzindo a neutralidade robótica.

3.3. Mapeamento de Expressões Não Manuais (ENMs)

Os rótulos emocionais são convertidos em animação facial via módulo procedural no Unity, utilizando blendshapes para deformação da malha. Cada estado afetivo possui um perfil de ativação aplicado de forma sobreposta e sincronizada à sinalização manual, preservando a gramática da Libras e adicionando a camada emotiva.

4. Avaliação Experimental e Resultados

A validação da arquitetura focou na precisão da detecção emocional e na superação da neutralidade expressiva, comparando o DeepSeek-V3 com abordagens tradicionais e mo-

delos de larga escala. O benchmarking utilizou um corpus de 300 enunciados do dataset GoEmotions. O protocolo confrontou o desempenho do DeepSeek-V3 com os sistemas GPT-4o-mini, Claude 3.5 Sonnet e Gemini 1.5 Flash.

4.1. Benchmarking de Modelos de Linguagem

O fine-tuning do BERT apresentou limitações (F1-score < 0,70) devido à sensibilidade ao ruído e falhas na detecção de nuances como ironia. Em contrapartida, o uso de LLMs via Prompt Engineering superou imediatamente esse patamar (ver Tabela 1). Embora o GoEmotions não seja específico para Libras, ele valida a inferência afetiva no texto em português, etapa que precede a tradução automática.

Tabela 1. Desempenho comparativo baseado em 300 enunciados do GoEmotions

Modelo Testado	F1-Score (Global)	Acurácia (Global)	Precisão em Raiva	Diagnóstico Técnico
DeepSeek-V3	0.90	90.0%	0.88 (Elevada)	Ideal. Equilibrado e seguro.
GPT-4o-mini	0.79	84.3%	0.59 (Reduzida)	Tende a alucinar hostilidade.
Claude 3.5 Sonnet	0.77	84.7%	0.57 (Reduzida)	Baixa especificidade com tendência a falsos positivos.
Gemini 1.5 Flash	0.73	73.0%	0.55 (Reduzida)	Confusão entre classes.

A categoria "Raiva"destaca-se no diagnóstico técnico por representar o maior desafio de classificação e risco à segurança comunicacional. Devido ao "viés de negatividade"em modelos menores, a precisão nesta classe indica a capacidade do sistema em distinguir nuances pragmáticas, evitando que expressões hostis indevidas comprometam a tradução para o usuário surdo.

4.2. Análise Visual da Prova de Conceito

Para validar a integração entre o motor de inferência e o módulo de renderização, foram realizados testes de transição emocional no Unity. A Figura 2 apresenta o avatar do VLibras reagindo a diferentes estados afetivos processados.

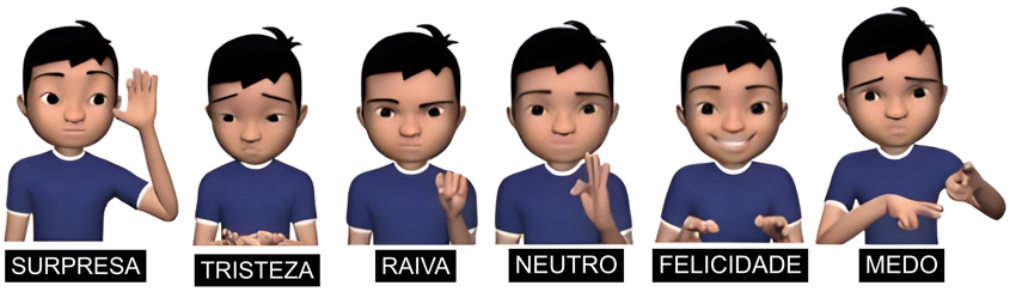

Figura 2. Expressões faciais no avatar a partir dos estados afetivos detectados

Nota-se que a deformação da malha facial via Blendshapes ocorre de forma sobreposta aos sinais manuais, preservando a gramática da Libras enquanto adiciona a camada emotiva. O sistema foi capaz de cobrir o espectro de emoções básicas (Felicidade, Tristeza, Medo, Surpresa e Raiva) em contraste com a expressão neutra padrão.

4.3. Discussão e Segurança Comunicacional

Os modelos concorrentes apresentaram "Viés de Negatividade"e baixa precisão em raiva, elevando o risco de interpretações pragmáticas errôneas. O DeepSeek-V3 mitigou esse cenário (0,88 de precisão), ampliando a confiabilidade do sistema. Para controlar alucinações, a arquitetura opera em três níveis: (1) restrição taxonômica via JSON; (2) *Prompt Engineering* com negativos; e (3) Inferência Dual, onde o contexto global garante a consistência das microexpressões.

A Inferência Dual permitiu identificar transições emotivas em períodos complexos (ex: "Tristeza"para "Felicidade"), superando a rigidez de sistemas tradicionais. A baixa latência é garantida pelo processamento assíncrono em Node.js e pela segmentação via *PhraseBreaker*, que viabiliza a interpolação procedural de *blendshapes* no Unity em sincronia com a sinalização manual, assegurando fluidez visual e robustez.

4.4. Ameaças à Validade e Estratégias de Mitigação

Para mitigar riscos de validade externa, o *benchmarking* foca na inferência do português pré-tradução. Quanto à validade de conclusão, o "Viés de Negatividade"e alucinações de hostilidade são combatidos via DeepSeek-V3 e *Prompt Engineering*, garantindo precisão na categoria "Raiva"e preservando a segurança comunicacional.

A validade de construto em frases complexas é resguardada pela Estratégia de Inferência Dual, que permite o acompanhamento dinâmico da intenção do discurso. Já o risco do "Vale da Estranheza"será mitigado em experimentos futuros com voluntários surdos, etapa essencial para validar a aceitabilidade social da tecnologia e refinar, via feedback direto, a suavização das transições faciais no Unity.

5. Conclusão

Este trabalho apresentou uma arquitetura modular integrada ao VLibras para mitigar a neutralidade expressiva na acessibilidade digital de surdos. Com o uso do modelo DeepSeek-V3, a pesquisa demonstrou a viabilidade inicial de uma inferência emotiva mais granular, apontando ganhos em relação a limitações observadas em abordagens tradicionais na identificação de estados emocionais.

A prova de conceito valida a integração entre o motor de inferência e a renderização no Unity, demonstrando que o sistema captura dinâmicas emocionais sequenciais e as converte em ENMs fundamentadas. O benchmarking e a validação indicam vantagens na mitigação do "Vale da Estranheza", ampliando a naturalidade da comunicação assistiva.

Como trabalhos futuros, o cronograma prevê a realização de experimentos controlados de usabilidade com voluntários surdos, essenciais para validar a percepção de naturalidade e aceitabilidade sob a ótica do público-alvo. Além disso, pretende-se refinar a suavização das transições faciais, assegurando que a tecnologia não apenas traduza palavras, mas transmita fielmente a intenção e o tom do discurso original.

Referências

Clemente, S. (2024). Análise das expressões faciais utilizadas na construção de sentenças interrogativas da libras. *Revista de Estudos Linguísticos*, 12(1).

Fang, S. et al. (2024). Signllm: Sign language production large language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2345–2356. IEEE.

Moraes, L. M. et al. (2018). A usabilidade de avatares de libras em sites: análise da interação de usuários surdos por meio do rastreador ocular eye tracking. *Design e Tecnologia*, 8(16):41–51.

Saunders, B., Camgoz, N. C., and Bowden, R. (2020). Progressive transformers for end-to-end sign language production. In *European Conference on Computer Vision (ECCV)*, pages 687–704. Springer.

Silva, V. M. (2021). O uso da análise de emoções como auxílio na tradução automática do português brasileiro para libras. Dissertação de mestrado, Universidade Federal da Paraíba.

Wolfe, R. and McDonald, J. C. (2021). A survey of facial nonmanual signals portrayed by avatar. *Grazer Linguistische Studien*, 93:161–223.