

LínguaQuiz: Integração de Reconhecimento de Imagens e LLM em um Jogo para Terapia de Fala

LínguaQuiz: Integration of Image Recognition and LLM in a Game for Speech Therapy

Pedro Henrique Brandão Oliveira Bastos¹, Victor Travassos Sarinho¹

¹Laboratório de Entretenimento Digital Aplicado (LEnDA)
Universidade Estadual de Feira de Santana (UEFS)
Feira de Santana, BA - Brasil

phbrandao@outlook.com, vsarinho@uefs.br

Abstract. Introduction: The use of digital games as a complementary tool in therapeutic processes has shown promise, especially in the field of speech therapy. **Objective:** This work presents the development of the game LínguaQuiz, designed to assist in the rehabilitation of oral movements through image capture and analysis using a Large Language Model (LLM). **Methodology or Steps:** The game uses the device's camera to capture images of the user's tongue movements, which are processed by a visual LLM. Depending on the tongue movements performed by the player, actions are interpreted and corresponding commands are sent to a quiz-style game, allowing therapy sessions to be conducted in a gamified format. **Results:** It is expected that LínguaQuiz will help make speech therapy rehabilitation exercises more dynamic, motivating, and effective, expanding the possibilities of support for clinical work.

Keywords Digital game, Image recognition, Speech therapy rehabilitation, Visual LLM, Gametherapy.

Resumo. Introdução: O uso de jogos digitais como ferramenta complementar em processos terapêuticos tem se mostrado promissor, especialmente no campo da fonoaudiologia. **Objetivo:** Este trabalho apresenta o desenvolvimento do jogo LínguaQuiz, projetado para auxiliar na reabilitação de movimentos orais por meio da captura e análise de imagens usando um Modelo de Linguagem de Grande Escala (LLM). **Metodologia ou Etapas:** O jogo utiliza a câmera do dispositivo para capturar imagens dos movimentos da língua realizados pelo usuário, que são processadas por um modelo de LLM visual. A depender dos movimentos da língua realizados pelo jogador, ações do jogador são interpretadas e respectivos comandos são enviados para um jogo estilo quiz, permitindo que as sessões terapêuticas possam ser realizadas de forma gamificada. **Resultados:** Espera-se que o LínguaQuiz contribua para tornar os exercícios de reabilitação fonoaudiológica mais dinâmicos, motivadores e eficazes, ampliando as possibilidades de apoio ao trabalho clínico. **Palavras-Chave** Jogo digital, Reconhecimento de imagens, Reabilitação fonoaudiológica, LLM visual, Gameterapia.

1. Introdução

Nos últimos anos, os métodos terapêuticos vêm incorporando, de forma crescente, tecnologias que proporcionam tratamentos mais interativos e envolventes. Com o aumento das expectativas dos pacientes, os profissionais de saúde têm se adaptado a essa tendência, adotando ferramentas computacionais como aplicativos de visualização, jogos e simuladores virtuais. Essas soluções oferecem abordagens mais abrangentes e menos convencionais, tornando os tratamentos mais agradáveis e eficazes [Sousa et al. 2011].

De acordo com [Weiss et al. 2004], na área da reabilitação, os jogos sérios possuem dois objetivos principais: ampliar a capacidade funcional dos pacientes e estimular sua participação na vida comunitária. Esses objetivos são alcançados tanto por meio da melhora de funções sensoriais, motoras e cognitivas, através de intervenções intensivas, quanto pela prática de atividades cotidianas e ocupacionais. Ainda segundo [Weiss et al. 2004], a intervenção terapêutica tem início com a execução de uma série de exercícios e atividades rotineiras, durante as quais as habilidades motoras e cognitivas dos pacientes são avaliadas. Essa avaliação contínua permite que a terapia seja constantemente ajustada, atendendo às necessidades específicas de cada indivíduo.

Os jogos representam um modo efetivo para fortalecer programas de reabilitação, não apenas pelo aumento do efeito motivacional, mas também porque desencadeiam um sistema de aprendizado mais eficiente baseado no reforço positivo que é fornecido quando o movimento correto é executado. Esse sistema capaz de fornecer recompensas melhora a retenção de ações que garantem a recompensa, facilita modificações no planejamento das ações e a generalização das habilidades [Schmid 2009].

Considerando o cenário dos jogos digitais aplicados à saúde e a carência de soluções acessíveis na área da fonoaudiologia, este estudo propõe o projeto e desenvolvimento do LínguaQuiz — um jogo digital de perguntas e respostas que utiliza reconhecimento de imagens através de um Modelo de Linguagem de Grande Escala (LLM), com o objetivo de atuar como ferramenta de apoio em sessões fonoaudiológicas, contribuindo para o processo de reabilitação dos pacientes.

2. Metodologia

A solução proposta consiste em uma interface para um jogo de perguntas e respostas, inicialmente de conhecimentos gerais, porém a temática das perguntas pode ser ajustada conforme a necessidade. Para simplificar o processo de detecção de imagens, as únicas respostas possíveis são “sim” ou “não”, que correspondem a movimentos da língua dos jogadores.

O jogo LínguaQuiz apresenta na tela uma sequência de perguntas com respostas possíveis de “sim” ou “não”. Durante cada partida, o jogador utiliza uma webcam para responder a essas perguntas, por meio da captura de uma imagem de seu rosto, acionada ao pressionar a tecla “P”. Para responder “sim”, o usuário deve capturar uma imagem exibindo a língua para fora da boca, em qualquer direção. Para responder “não”, basta capturar uma imagem com a boca fechada. Após cada envio, o jogo exibe uma mensagem informando se a resposta do jogador estava correta ou incorreta.

Para facilitar a detecção e interpretação das imagens, o LínguaQuiz utiliza a LLM

```
async function run() {
  const prompt = "Is there a person showing the tongue? Answer just 'yes.' or 'no.'";
  const imageEncoded = {
    inlineData: {
      data: Buffer.from(fs.readFileSync("image.jpg")).toString("base64"),
      mimeType: "image/jpeg"
    }
  }

  const result = await model.generateContent([prompt, imageEncoded], safe);
  const response = result.response.text();

  const text = response.replace(/[\.\s]/g, '').toLowerCase();
  console.log(text);

  res.send({ data: text });
}
```

Figura 1. Processamento de prompt em linguagem natural através da API do Google Gemini. Fonte: Autoria Própria

Google Gemini¹, que conta com modelos avançados de reconhecimento de imagens. Assim, as imagens capturadas da língua do jogador são enviadas para a API do Gemini, que identifica se o usuário está ou não com a língua para fora e, com base nisso, executa as ações correspondentes no jogo. A escolha pelo Google Gemini se deu pela facilidade de integração e pela disponibilidade de uma cota gratuita, embora limitada a 5 solicitações por minuto — o que exigiu a inclusão de intervalos de tempo entre as perguntas para evitar o excesso de requisições.

O jogo adota uma arquitetura cliente-servidor, onde um aplicativo no navegador envia imagens para um servidor central. Esse servidor, desenvolvido em Node.js², recebe a imagem e a transforma em um prompt de linguagem natural, solicitando ao Google Gemini que informe se a língua está visível ou não. A resposta retornada pela API é então enviada de volta ao cliente, que dá continuidade à partida de acordo com o resultado. A Figura 1 ilustra um trecho do código do servidor responsável por processar o prompt e a imagem enviada, utilizando a API do Google Gemini.

No lado do cliente, foi implementado um jogo de perguntas e respostas no qual o jogador interage por meio de uma webcam, que captura imagens de seu rosto conforme as instruções apresentadas na tela. Para isso, foi utilizado o Next.js³, um framework para o desenvolvimento de aplicações web que emprega a biblioteca React⁴ na construção de componentes reutilizáveis ao longo do projeto. A funcionalidade de captura da webcam foi desenvolvida com a biblioteca React Webcam⁵, que disponibiliza recursos de exibição em tempo real da câmera do usuário e captura de imagens estáticas. Por fim, o projeto foi estilizado com o framework Tailwind CSS⁶, que oferece uma coleção abrangente de classes utilitárias em CSS, permitindo a criação de interfaces personalizadas de forma ágil e flexível.

¹<https://gemini.google.com>

²<https://nodejs.org>

³<https://nextjs.org/>

⁴<https://react.dev/>

⁵<https://www.npmjs.com/package/react-webcam>

⁶<https://tailwindcss.com/>

3. Jogo Obtido

Ao iniciar o LínguaQuiz, o jogador se depara com a tela inicial, onde são apresentadas a interface do jogo e as instruções para a interação. No canto superior esquerdo dessa tela, localiza-se o módulo da câmera, destacado em vermelho na Figura 2. Por meio desse módulo, o jogador tem acesso a uma imagem contínua de seu rosto, que serve como referência visual para auxiliar na tomada de decisões durante a partida. Para isso, é necessário que o dispositivo possua uma webcam instalada e com as permissões de uso devidamente configuradas. A imagem exibida no momento em que o comando para capturar a foto for acionado será enviada ao servidor, que processará a imagem e retornará a resposta correspondente ao cliente.

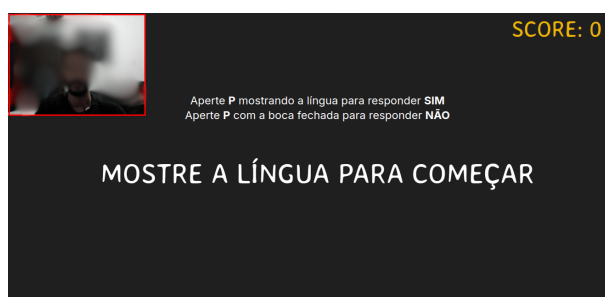


Figura 2. Tela inicial do jogo LínguaQuiz. Fonte: Autoria Própria.

No centro da tela (Figura 2), encontram-se as instruções sobre como jogar o LínguaQuiz. Conforme as perguntas aparecem, o jogador deve pressionar a tecla “P” (de *print*, ou captura) para enviar sua resposta, que pode ser “sim” ou “não”. Para responder “sim”, o jogador deve projetar a língua para fora da boca e pressionar o comando. Para responder “não”, basta manter a língua dentro da boca e pressionar a mesma tecla. Antes do início da partida, é exibida a mensagem “Mostre sua língua para começar”, sinalizando que o jogador precisa realizar esse movimento para iniciar o jogo. As instruções visuais para a interação correta estão apresentadas na Figura 3.

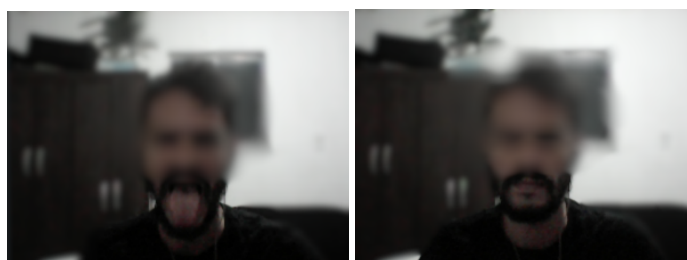


Figura 3. Interação com o jogo para registrar as respostas “sim” e “não”. Fonte: Autoria Própria.

No canto superior direito da tela inicial, conforme destacado na Figura 2, encontra-se a área de pontuação, identificada como “Score”. Antes de iniciar o jogo, a pontuação é zerada. Durante a partida, a cada resposta correta, o jogador acumula 20 pontos, enquanto respostas incorretas não somam pontos. Quando o jogo é iniciado, uma pergunta é exibida na tela — como mostrado na Figura 4, destacada em vermelho. As perguntas são pré-configuradas e não podem ser alteradas diretamente pela interface do jogo; qualquer modificação só é possível editando manualmente um arquivo no diretório

do projeto. O jogador pode então escolher entre duas ações: responder “sim” ou “não”. Caso a resposta seja “sim”, e esteja correta, o jogo exibirá uma mensagem de sucesso. Se a resposta for “não”, e incorreta, será exibida uma mensagem de erro. Em ambos os casos, um estímulo sonoro é reproduzido junto com o aviso visual, proporcionando maior imersão ao jogador.

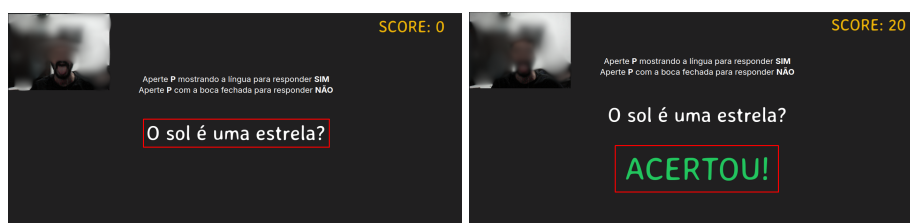


Figura 4. Exemplo de uma pergunta do jogo, com resposta correta. Fonte: Autoria Própria.

Após a resposta de todas as perguntas configuradas, o jogo exibe na tela a pontuação final do jogador. Em seguida, após alguns segundos, o sistema retorna automaticamente para a tela inicial, permitindo o início de uma nova partida.

O protótipo desenvolvido⁷ foi disponibilizado para testes na plataforma Vercel⁸, que oferece infraestrutura em nuvem e ferramentas gratuitas para desenvolvimento, hospedagem e escalonamento de aplicações web. Já o servidor Node.js responsável pelo processamento das imagens foi hospedado no Render⁹, serviço que permite aos desenvolvedores implantar e escalar aplicações de maneira simples e eficiente.

4. Conclusões e Trabalhos Futuros

Este trabalho apresentou o LínguaQuiz, um jogo digital estilo quiz que utiliza o reconhecimento visual dos movimentos da língua como interface de interação com o usuário. Trata-se de uma ferramenta que pode ser empregada por fonoaudiólogos em terapias voltadas para a reabilitação dos movimentos da língua, em especial para pacientes que mostram dificuldade em manter a língua dentro da boca com os lábios fechados.

Baseado em um modelo de perguntas e respostas, o LínguaQuiz demonstrou-se eficaz para estimular a coordenação dos músculos da língua — habilidade essencial para a articulação correta da fala. Os resultados obtidos neste estudo evidenciam a importância do desenvolvimento de jogos digitais que utilizam o corpo como interface, aplicados ao processo de reabilitação da fala e da linguagem. Além disso, a personalização e os elementos de gamificação empregados na terapia fonoaudiológica oferecem novas possibilidades para tornar as sessões terapêuticas mais motivadoras, dinâmicas e eficazes para os pacientes.

Como proposta para trabalhos futuros, pretende-se criar jogos mais elaborados que utilizem os movimentos da língua, como jogos de plataforma e de corrida. Além disso, planeja-se aplicar essa abordagem de terapia gamificada em diferentes perfis de usuários, considerando fatores como faixa etária, condições de saúde e necessidades específicas de reabilitação.

⁷<https://LinguaQuiz.vercel.app>

⁸<https://vercel.com/>

⁹<https://render.com/>

Referências

- Schmid, M. (2009). Reinforcing motor re-training and rehabilitation through games: a machine-learning perspective. *Frontiers in neuroengineering*, 2:1016.
- Sousa, A. d. S. et al. (2011). Serious games para a fonoaudiologia: uma abordagem voltada à terapia em motricidade orofacial.
- Weiss, P. L., Rand, D., Katz, N., e Kizony, R. (2004). Video capture virtual reality as a flexible and effective rehabilitation tool. *Journal of neuroengineering and rehabilitation*, 1:1–12.