

# Aplicando o Reconhecimento Visual de Movimentos da Língua para Dinâmicas de Jogos Digitais em Terapias Fonoaudiológicas

Lenington C. Rios<sup>1</sup>, Washington P. Batista<sup>1</sup>, Victor T. Sarinho<sup>1</sup>

<sup>1</sup>Laboratório de Entretenimento Digital Aplicado (LEnDA)  
UEFS - Feira de Santana - Bahia - Brasil

{lenitorios, wstrokes}@gmail.com, vsarinho@uefs.br

**Abstract.** *Speech therapies are often repetitive and tiring, causing impatience and often the treatment discontinuation by children and adolescents. This work presents the application of visual recognition techniques in the interpretation and capture of tongue movements performed in a speech therapy. As a result, it is possible to integrate basic exercises of oral muscle strengthening with several digital game dynamics, allowing a better acceptance of the patients to the routines applied in the respective therapies.*

**Resumo.** *Terapias fonoaudiológicas costumam ser repetitivas e cansativas, causando impaciência e por muitas vezes a descontinuidade do tratamento de crianças e adolescentes. Este trabalho apresenta a aplicação de técnicas de reconhecimento visual na interpretação e captura de movimentos da língua executados em terapias fonoaudiológicas. Como resultado, tem-se a possibilidade de integrar exercícios básicos de fortalecimento muscular oral com dinâmicas de jogos digitais diversos, permitindo assim uma maior aceitação dos pacientes as rotinas aplicadas nas respectivas terapias.*

## 1. Introdução

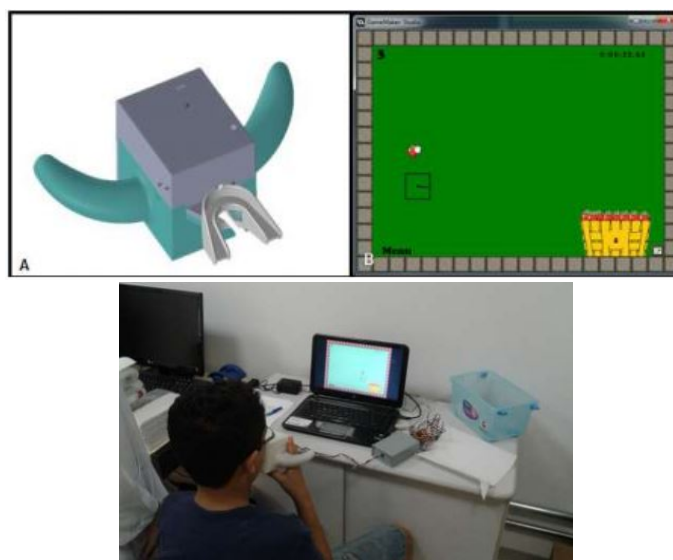
Os avanços tecnológicos dos últimos anos trouxeram benefícios para profissionais da área de saúde, auxiliando-os na definição de diagnósticos precisos, na intervenção adequada de pacientes e na execução de atividades de reabilitação de uma maneira prazerosa e interativa [Sousa et al. 2011]. Tais avanços contribuem em “larga escala para a solução de problemas antes insolúveis e que podem reverter em melhores condições de vida e saúde para o paciente” [Barra et al. 2006].

Uma das áreas tecnológicas que vem ganhando destaque é a dos *serious games*, que são aplicações que se estendem além dos jogos tradicionais, incluindo a educação, gestão pública, treinamento militar, cuidados de saúde e terapia, dentre outros [Orozco et al. 2012]. Tecnologias de computação visual também possibilitam ao usuário uma interação diferenciada com o computador, além de oferecer uma forma mais eficiente de envolvimento do usuário com os sistemas de software em si [Rodrigues et al. 2011].

Este artigo apresenta uma proposta de serious game para aplicação de exercícios de musculatura oral utilizados em terapias fonoaudiológicas via tecnologia de reconhecimento visual. Tal proposta tem como princípio a captura de movimentos da língua do paciente (esquerda e direita inicialmente) para serem aplicados em mini-jogos terapêuticos diversos.

## 2. Trabalhos Relacionados

Como exemplo de trabalho relacionado ao uso de jogos digitais em terapias fonoaudiológicas, Amaral et. al [Amaral et al. 2015] desenvolveu um joystick para a reabilitação da força da língua. Trata-se do uso de um instrumento bucal para o encaixe na cavidade oral que possui uma peça de comando para a língua (Figura 3). Esta peça pode ser movida para direita, esquerda, cima ou para baixo, a qual interage com uma interface desenvolvida para um jogo [Amaral et al. 2015]. O jogo consiste de alvos (frutas) que devem ser alcançados pelo usuário após aplicar uma determinada força no controle com a língua. Com o término de uma partida, o jogo informa o desempenho final do paciente, apresentando dados do paciente, dados do jogo, e dados do desempenho do jogador, tais como: número de tentativas, força máxima e média aplicada, tempo máximo e médio de contração, alvos alcançados, área e tempo necessários para alcançar o alvo no jogo [Amaral et al. 2015].



**Figura 1. Instrumento para capturar o movimento da língua, interface do jogo e aplicação do mesmo em um paciente [Amaral et al. 2015].**

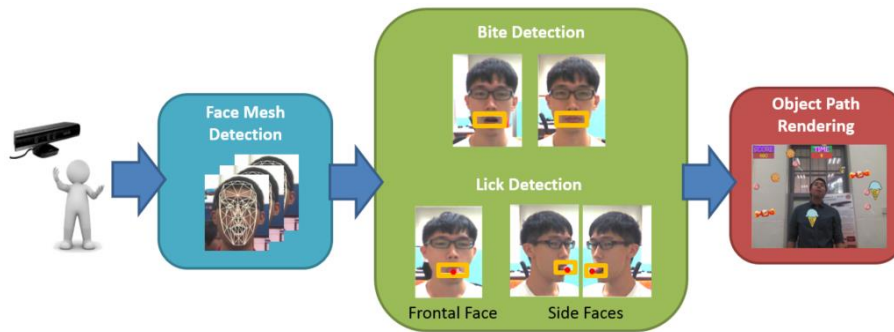
Kothari et al. [Kothari et al. 2014] introduziu um jogo de computador como método de treinamento para pacientes com incapacidade motora na língua através de um *Tongue Drive System* (TDS), uma tecnologia assistiva inicialmente desenvolvida para pacientes com quadriplegia. TDS permite que pacientes tenham controle sobre um computador, uma cadeira de rodas elétrica ou o ambiente do usuário usando a língua como um manipulador. Como resultado, o ato de jogar o jogo de computador com a língua, com ênfase no treinamento de habilidades motoras, induziu a plasticidade cortical no corticômotor humano de controle da língua, algo que permite aplicar tal abordagem como um novo regime adjuvante de reabilitação neurológica no tratamento de pacientes com deficiência de motricidade na língua.

Para finalizar, Pan et al. [Pan et al. 2015] apresentou um sistema interativo de reabilitação oral, que detecta mordidas ou lambidas com base nos movimentos da boca e da língua com o auxílio do Kinect. A fim de realizar fisioterapia oral, esse sistema incita o usuário a exercitar seus músculos orais jogando jogos. O usuário recebe pontuações



**Figura 2. Participante vestindo touca do TDS com controle do cursor do computador através do imã ligado à língua [Kothari et al. 2014].**

quando ele morder ou lamber os alimentos virtuais que caem do céu. Os movimentos dos lábios são medidos para detectar a ação da mordida, e diferentes estratégias são projetadas para localizar a ponta da língua para várias orientações da face, de modo que a ação da lambida possa ser detectada com mais precisão.



**Figura 3. Framework proposto para o sistema de detecção de mordidas e lambidas com a língua [Pan et al. 2015].**

### 3. Metodologia e Resultados

Para realizar a captura de movimentos da língua (caracterizado, nessa proposta de trabalho, como a posição final da língua para a esquerda, direita ou nenhum desses movimentos), realizou-se um treinamento para a detecção de imagens da boca e da língua através da biblioteca *TensorFlow*. Trata-se de uma máquina de aprendizagem *open source* desenvolvida pela *Google* para treinamentos, aquisição de dados, entre outros [Abadi et al. 2016]. Essa biblioteca deriva de operações matemáticas realizadas pelas redes neurais através de dados multidimensionais, representados por matrizes que são conhecidos como tensores [Do et al. 2017] [TensorFlow 2018].

Para o treinamento do TensorFlow, efetuou-se uma coleta de imagens da boca de um total de 28 pessoas voluntárias ao projeto. Dessas pessoas, foram obtidas 170 imagens, sendo: 72 fotos com a língua para a esquerda; 100 com a língua para a direita; e 50 com a boca fechada. Em seguida, realizou-se a extração de características de cada

foto utilizando o *TensorFlow* para então obter o classificador para detecção da posição da língua.

Utilizando uma *real time tracking* através da *webcam*, realizou-se a execução do classificador treinado para a obtenção de imagens e a análise de seus resultados. O *real time tracking* consiste na captura de imagens da câmera, ou seja, quando acontece o movimento da língua (para esquerda, direita ou nenhum dos dois) é desenhado um quadrado e apresentado qual movimento foi realizado, com sua relativa porcentagem de precisão.

Inicialmente, foram realizados testes com imagens de uma boca com movimentos da língua para a direita e para a esquerda. Como resultado, a Figura 4 mostra os níveis de detecção obtidos com o treinamento, demonstrando precisões aproximadas de 50% para a língua na posição direita, 57% para a língua na posição esquerda e 99,8% para a boca fechada.



**Figura 4. Precisão obtida com o classificador para detecção da língua nas posições direita, esquerda e de boca fechada.**

A Tabela 1 também mostra os resultados de precisão obtidos com as imagens testes para o classificador. Nela, é possível observar resultados positivos do classificador indicando a posição correta da língua, sendo “IMAGEM 1” com precisão de 61% para a direita e “IMAGEM 2” com precisão de 62% para esquerda.

**Tabela 1. Resultados do classificador com imagens teste da língua para esquerda e direita usando TensorFlow.**

Precisão	IMAGEM 1	IMAGEM 2
Esquerda	0.39281	0.62447
Direita	0.60542	0.32327
Boca fechada	0.00177	0.05226

Com relação ao uso do classificador produzido em dinâmicas de jogos digitais, utilizou-se a ferramenta *Godot 2.1.3* [Linietsky et al. 2014] para construir um mini-jogo capaz de utilizar os movimentos capturados da língua do jogador. A ideia é que a posição da língua atue como uma forma de controle dos movimentos da língua de um personagem do jogo, neste um sapo esfomeado que quer comer todas as moscas que aparecem no jogo.

Para tal, a posição da língua do jogador é capturada pela *webcam*, a qual, após ser interpretada pelo classificador *Tensorflow*, é fornecida para o jogo em execução através da leitura contínua de um *buffer* de dados compartilhado. O jogo lê este *buffer*, fica ciente da posição corrente da língua e, assim que o usuário pressionar o botão “OK” no jogo (Figura 5), a língua do sapo irá acompanhar a posição da língua do jogador. Caso uma mosca esteja no caminho da língua, o sapo irá comê-la, marcando assim pontos para o paciente jogador.



Figura 5. Imagem do mini-jogo em produção com a ferramenta Godot.

#### 4. Conclusões e Trabalhos Futuros

Este artigo apresentou o uso de tecnologias de reconhecimento visual para desenvolvimento de classificadores de movimentos da língua. Para tal, foi descrita uma abordagem de treinamento aplicada utilizando o TensorFlow para a configuração de classificadores de reconhecimento, bem como o estado atual de desenvolvimento de um mini-jogo capaz de utilizar o classificador escolhido em dinâmicas de jogos digitais terapêuticos.

Dentre os resultados obtidos, comprovou-se a viabilidade da integração de tecnologias de reconhecimento visual com dinâmicas de jogos digitais capazes de serem usados em exercícios orais de terapias fonoaudiológicas. O uso de reconhecimento visual também se mostrou capaz de evitar a aplicação de uma interface bucal física para controles de língua, a qual poderia gerar algum tipo de desconforto em seu uso, especialmente para pacientes portadores de necessidades especiais. Contudo, como relatado previamente, ainda se faz necessário ampliar o nível de precisão no reconhecimento visual do classificador treinado, de modo a permitir a aplicação do mesmo em ambiente real de terapia com crianças e adolescentes.

Como trabalhos futuros, pretende-se melhorar a eficiência de detecção do classificador da língua, visto que ainda existe o incômodo pela busca de uma posição frente à *webcam* para obter o resultado procurado. A captura de movimentos da língua no eixo vertical, bem como a detecção de movimentos outros tipos de exercícios orofaciais também

se encontra em estudo. Para finalizar, o desenvolvimento de uma coletânea de mini-jogos terapêuticos capazes de explorar a integração com o classificador treinado também será abordado em um futuro próximo.

## Referências

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: a system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283.
- Amaral, M. S., Moraes, R. M. M., Furlan, E. B. d. L. C., and Motta, A. R. (2015). Utilização de método para reabilitação da força da língua associado a jogos computacionais: estudo piloto. *XXII Congresso Brasileiro e IX Congresso Internacional de Fonoaudiologia*.
- Barra, D. C. C., Nascimento, E. R. P. d., Martins, J. d. J., Albuquerque, G. L., and Erdmann, A. L. (2006). Evolução histórica e impacto da tecnologia na área da saúde e da enfermagem. *Revista Eletrônica de Enfermagem*, 8(03):422–430.
- Do, Q., Son, T. C., and Chaudri, J. (2017). Classification of asthma severity and medication using tensorflow and multilevel databases. *Procedia Computer Science*, 113:344–351.
- Kothari, M., Svensson, P., Jensen, J., Holm, T. D., Nielsen, M. S., Mosegaard, T., Nielsen, J. F., Ghovanloo, M., and Baad-Hansen, L. (2014). Tongue-controlled computer game: a new approach for rehabilitation of tongue motor function. *Archives of physical medicine and rehabilitation*, 95(3):524–530.
- Linietsky, J., Manzur, A., and Community, G. (2014). Godot docs. <http://docs.godotengine.org/>. Accessed: 2019-04-20.
- Orozco, M., Silva, J., El Saddik, A., and Petriu, E. (2012). The role of haptics in games. In *Haptics rendering and applications*. InTech.
- Pan, T.-Y., Wong, Y.-X., Lee, T.-C., and Hu, M.-C. (2015). A kinect-based oral rehabilitation system. In *Orange Technologies (ICOT), 2015 International Conference on*, pages 71–74. IEEE.
- Rodrigues, H. F. et al. (2011). Aplicando sistemas hápticos em serious games: um jogo para a educação em higiene bucal. *Universidade Federal da Paraíba*.
- Sousa, A. d. S. et al. (2011). Serious games para a fonoaudiologia: uma abordagem voltada à terapia em motricidade orofacial. *Universidade Federal da Paraíba*.
- TensorFlow (2018). An open source machine learning library for research and production. <https://www.tensorflow.org/?hl=pt-br>. Accessed: 2019-04-20.